University of Alberta

Multiple Myeloma with t(4;14)(p16;q32): Biology and Clinical Impact

by

James "Jonathan" Herbert Keats

©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Department of Oncology

Edmonton, Alberta

Fall 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08663-7

Your file Votre référence
ISBN:
Our file Notre reférence
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Dedication

In memory of Peter Franklin, Phil Stepney and all those who contributed to our research program with the hope of a better outcome. Their dedication and fortitude should never be forgotten.

Abstract

Multiple Myeloma (MM) is currently an incurable cancer associated with a median survival of 4.5 years. This highly aggressive disease is clinically associated with a plasma cell infiltration of the bone marrow, a serum monoclonal protein spike, and is commonly associated with lytic bone lesions. The most common genetic abnormality in this cancer is a variety of different translocations involving the immunoglobulin heavy chain (IgH) locus. In general, these IgH translocations are presumed to be initiating events in the pathogenesis of MM as genes located on partner chromosomes are overexpressed due to their proximity with IgH enhancer elements. The second most common IgH translocation in MM is t(4;14)(p16;q32). This translocation typically occurs in the switch regions of the IgH locus, thus they are proposed to result from errors in the class switch recombination process. Since the breakpoints are in switch regions, the enhancer regions of the IgH locus are separated onto different derivative chromosomes resulting in dysregulted gene expression on both derivatives. We identified this translocation in 14.4% of patients within our MM patient cohort and showed that it predicted for a poor overall outcome. Suprisingly, outcome was independent of FGFR3 expression, the originally proposed t(4;14) target gene. The loss of FGFR3 expression was associated with the loss of the der(14) chromosome, suggesting t(4;14) is not always a reciprocal translocation. Subsequently, we identified three transcripts from the MMSET locus which are universally overexpressed in all t(4;14) positive patients. However, only transcripts encoding RE-IIBP produce protein products with uncompromised function in all t(4;14) positive patients. Therefore, the overexpression of RE-IIBP is likely essential to the establishment and maintenance of this highly aggressive subtype of MM. Finally, by cloning genomic translocation breakpoints, we identified two equally common translocation mechanisms. In 50% of patients, the translocation appears to originate from class switch recombination errors but this does not appear to be the case in the remaining patients. In summary, t(4;14) occurs by two equally common but different processes in MM, which result in the overexpression of RE-IIBP and is ultimately associated with a highly aggressive form of MM.

Acknowledgement

The body of work presented in this thesis is the cumulative effort of a great number of individuals. First and foremost I would like to thank my supervisor, Dr. Linda Pilarski, PhD for providing me with the opportunity to grow and develop, both scientifically and personally, within her research group. Under Linda's mentoring I've progressed from a timid graduate student, who forgot to breathe when he asked his first question at an international meeting, to one that considers himself a leader within his field. At a personal level the compassion and patience I've received from Linda in the last year has made a very trying experience much easier than I could ever put into words, but I'd just like to say thank you from the bottom of my heart. I've also had the great opportunity to develop an appreciation for the clinical side of multiple myeloma through our interaction with Dr. Andrew Belch, MD. Andy always provides our grounding influence and often our comedic relief.

I've been very fortunate to work with a number of individuals during the course of this work. First, I'd like to acknowledge Karen Seeberger for her initial contributions and work in developing the hybrid transcript assays within our lab. I often wonder how much of this project would have occurred, if it was not for Karen's initial work. I also need to thank my "work crew" of Jennifer Szydlowski and Darlene Paine, who helped organize the fridge and helped in the initial patient screening work. Erin Strachan and Tara Tiffinger have made significant contributions to much of my work and have more or less been acting as my big or little sisters. Moreover, most of the clinical data extraction and clinical analysis was done by Dr. Tony Reiman, MD. Tony has quickly become an integral member of our research group and a good friend. I'm indebted to Dr.

Christopher Maxwell, PhD my great friend and partner in crime through our graduate program. Chris did all of the immunoblotting experiments and helped with many of the imaging experiments, but more importantly he provided a sounding block for many good and bad ideas. Finally, much of this work has benefited from the help of Dr. Brian Taylor, PhD who is our go to resource within the lab.

I've been extremely fortunate to receive a helping hand from Dr. Leif Bergsagel, MD and Dr. Marta Chesi, PhD. They provided me with several plasmid constructs and cell lines which were essential to my research. I'm also indebted to Dr. Takemi Otsuki for so courteously providing so many cell lines.

Lastly, I need to thank my family and friends for all of their support over the years and in particular the last year. I would never have survived many of the difficult and trying days, if it was not for their patience and compassion.

Preface

A great deal of this thesis incorporates data generated by a variety of genome sequencing projects; specifically the *Homo sapiens, Mus musculus, Rattus norvegicus, Canis familiaris, Pan troglodytes, Gallus gallus, Xenopus tropicalis, Tetraodon nigroviridis, Takifugu rubripes, Danio rerio, Apis mellifera*, and *Drosophila melanogaster* genome projects. In an attempt to simplify the nomenclature I have used the proper gene names accepted by the Human Genome Organisation (HUGO) nomenclature committee (www.gene.ucl.ac.uk/nomenclature/). The one exception is the use of *MMSET* as opposed to *WHSC1* in most circumstances, in keeping with the typical gene nomenclature used within the multiple myeloma field. Other rare exceptions are present in the evolutionary analysis where proper names have not been assigned and thus the HUGO name of the nearest human homolog is used. In the case of the noted gene mutations, the nomenclature accepted by the Human Gene Mutation Database (HGMD)(www.hgmd.org/) is followed. Furthermore, mutation positions are reported using the reference sequence used by HMGD and any variations are noted in attached footnotes.

Table of Contents

Introduction 1

I.1.1 – The Human Face of Multiple Myeloma 2

I.2 – General Background on Multiple Myeloma 5

I.2.1 – Demographics in Canada 5

I.2.2 – Differential Diagnosis and Follow-up 5

I.2.3 – Phenotype and Hierarchy of Myeloma Cells 10

I.3 - The Underlying Genetics of Multiple Myeloma 14

I.3.1 – Common Oncogenes and Tumour Suppressors in Myeloma 14

I.3.2 – Ploidy and Myeloma 17

I.3.3 – Translocations in Myeloma 20

I.4 - The t(4;14)(p16;q32) Translocation 24

I.4.1 - The Initial Description of t(4;14)(p16;q32) in Multiple Myeloma 25

I.4.2 - Improving the Characterization of t(4;14)(p16;q32) 27

I.4.3 - The Incidence of t(4;14)(p16;q32) in Multiple Myeloma 37

I.4.4 - The Significance of t(4;14)(p16;q32) in Multiple Myeloma 41

I.5 - The Genes Associated with t(4;14)(p16;q32) 44

I.5.1 - TACC3/ERIC-1 46

I.5.2 - FGFR3/ACH/CEK2/JTK4/HSFGFR3EX 48

I.5.3 - LETM1 58

I.5.4 - WHSC1/MMSET/NSD2/TRX5 59

I.5.5 - WHSC2/NELF-A 72

Materials and Methods 74

MM.1 – Origin of Analyzed Samples 75
MM.1.1 - Cell Lines 75
MM.1.2 – Patient Samples 75
MM.1.3 – Purification of Bulk Samples 76
MM.1.4 – Plasma Cell Purification 77
MM.2.1 – Clinical Data Extraction 77
MM.2.2 – Clinical Statistics 78
MM.3 – Molecular Sample Isolation and Analysis 79
MM.3.1 - RNA and DNA Isolation 79
MM.3.2 – Synthesis of cDNA for RT-PCR Applications 79
MM.3.3 – Standard RT-PCR Reactions 79
MM.3.4 - Quantitative RT-PCR Reactions 82
MM.3.5 – PCR Amplification of Genomic t(4;14) Breakpoints 83
MM.4 – Analysis of MMSET Protein Variants 88
MM.4.1 - Expression Vector Construction 88
MM.4.2 - Transfections, Immunoblotting and Microscopy 89
Chapter 1: Frequency and Clinical Significance of t(4;14)(p16;q32) in Multiple Myeloma and Monoclonal Gammopathy of Undetermined Significance 92
C1.1.1 - Brief Introduction 93
C1.2.1 – Specific Aim 95
C1.3 – Chapter 1 Results 96
C1.3.1 – Sample Acquisition and Screening Assays 96
C1.3.2 – The Occurrence of t(4;14)(p16;q32) 102

C1.3.3 – The Clinical Impact of t(4;14)(p16;q32) 106

C1.3.4 – Monitoring Disease Burden with IgH-MMSET Hybrid Transcript Assays 110

C1.3.5 – Unique Observations from the der(4) IgH-MMSET Screening Assays... 113

Chapter 2: Target Genes and the der(14) Chromosome 119

C2.1.1 - Brief Introduction 120

C2.2.1 - Working Hypothesis 122

C2.3 – Chapter 2 Results 123

C2.3.1 – The Expression of *FGFR3* in t(4;14) Positive Patients 123

C2.3.2 – Clinical Outcome Associated with t(4;14) and/or *FGFR3* Expression.... 127

C2.3.3 – The der(14) Chromosome is Undetectable in *FGFR3* non-expressers 130

C2.3.4 – Exploratory Clinical Analysis 135

C2.3.3 – Alternative Target Genes 138

C2.3.6 – RE-IIBP is Overexpressed in BMMC of t(4;14) Positive Patients 146

C2.4.1 – Chapter Conclusions 149

Chapter 3: Analysis of MMSET Protein Variants Associated with t(4;14) Myeloma .. 152

C3.1.1 - Brief Introduction 153

C3.2.1 - Working Hypothesis 155

C3.3 – Chapter 3 Results 156

C3.3.1 – MMSET Constructs 156

C3.3.2 – Localization of Wild-type MMSET Protein Variants 159

C3.3.3 – Localization of Myeloma Breakpoint Specific MMSET Protein Variants 163

C3.3.4 – Characterization of a Novel Localization Domain within MMSET III ... 167

C3.3.5 – Functional Assessment of MMSET Protein Variants 170

C3.3.6 – Evolutionary Analysis of MMSET 172

C3.4.1 – Chapter Conclusions 190

Chapter 4: Analysis of t(4;14) Genomic Breakpoints 193

C4.1.1 - Brief Introduction 194

C4.2.1 – Specific Aim 196

C4.3 – Chapter 4 Results 197

C4.3.1 – The Breakpoint Cloning Strategy 197

C4.3.2 – The Cloned t(4;14)(p16;q32) Breakpoints 200

C4.3.3 – Modeling the Cloned t(4;14)(p16;q32) Breakpoints 203

C4.3.4 – Analysis of the Cloned t(4;14)(p16;q32) Breakpoints 213

C4.3.5 – Description of the t(4;14) Breakpoints Cloned from the Edmonton Cohort 221

C4.3.6 – Description of the t(4;14) Breakpoints Cloned from Myeloma Cell Lines 225

C4.4.1 – Chapter Conclusions 228

Discussion 230

D.1.1 – The Frequency of t(4;14)(p16;q32) in Myeloma and MGUS 231

D.1.2 – FGFR3 Expression and the der(14) Chromosome 236

D.1.3 – The Clinical Significance of t(4;14)(p16;q32) in Multiple Myeloma 240

D.1.4 – The Overexpression of MMSET Transcripts in t(4;14)(p16;q32) Myeloma 245

D.1.5 – The MMSET Protein Variants 248

D.2.1 – Model of t(4;14)(p16;q32) Mediated Myelomagenesis 252

D.3.1 – Significance of This Work 262

Bibliography 265

Appendix I: Detailed Screening Data From the t(4;14) Positive Patients 292

List of Tables

Table I.1 – Differential Diagnosis of Monoclonal Gammopathies 8

Table I.2 – The Incidence of t(4;14) in Multiple Myeloma Patients 38

Table I.3 – Known FGFR3 Mutations 53

Table MM.1 - Standard RT-PCR Primers 80

Table MM.2 – Quantitative RT-PCR Primers and Probes 83

Table MM.3 – PCR Primers used to Clone and Sequence t(4;14) Breakpoints 86

Table MM.4 – Primers Used to Clone MMSET Variants 88

Table C1.1 – The Incidence of t(4;14)(p16;q32) in Myeloma and MGUS 103

Table C1.2 – Clinical Characteristics of Patients in the Original Cohort 107

Table C2.1 – Frequency of *FGFR3* Expression in t(4;14) Positive Patients 127

Table C2.2 – Results of the der(14) MMSET-IgH Hybrid Transcript Assays 133

Table C2.3 – Correlation between Clinical Isotype and der(14) Result 134

Table C2.4 - Quantitative Relative Expression Level of Genes Located at 4p16.3 in Cell Lines 142

Table C2.5 – Quantitative Relative Expression of Potential t(4;14) Target Genes in *ex vivo* Plasma Cells 145

Table C3.1 – Features of MMSET Protein Variants 159

Table C3.2 – FRAP $t_{1/2}$ Times of the Various MMSET Variants 171

Table C4.1 – Attempted Breakpoint Cloning Results 202

Table C4.2 – List of Patients Fitting Each Translocation Model 211

Table C4.2 – Tabulated Breakpoint Information 216

Table D.1 – The Incidence of t(4;14) in MGUS 233

Table D.2 – The Distribution of t(4;14)(p16;q32) Breakpoint Types.............................. 235

Table D.3 – FGFR3 Expression in t(4;14) Positive Samples 239

List of Figures and Illustrations

Figure I.1 – Typical Plasma Cell Morphologies Observed in Myeloma Patients 11

Figure I.2 – Basic Diagram of t(4;14)(p16;q23) 25

Figure I.3 – FISH Strategies Used by Different Groups to Detect t(4;14) 32

Figure I.4 – Phylogeny of the FGFR, TACC, and NSD paralogs 45

Figure I.5 – Alternative Splicing and Initiation of Transcription of MMSET Results in Various Wild-type Proteins 63

Figure I.6 – The SET2 Histone Methyl-Transferase Family 71

Figure C1.1 – Test of Potential RT-PCR Based Screening Reactions 98

Figure C1.2 – Description of the der(4) IgH-MMSET Hybrid Transcript Assays 101

Figure C1.3 – Sensitivity of the der(4) IgH-MMSET Hybrid Transcript Assays 102

Figure C1.4 – Representative Panel of the t(4;14) Positive Patients 104

Figure C1.5 – Kaplan-Meier Overall Survival Plot for the Original Patient Cohort 109

Figure C1.6 – Timeline Analysis of IgH-MMSET Hybrid Transcripts 112

Figure C1.7 – Patient with both MB4-2 and Novel MB4-4 Hybrid Transcripts 116

Figure C2.1 – The Mechanism of *FGFR3* Dysregulation on the der(14) 125

Figure C2.2 – *FGFR3* Transcripts are not Detectable in all t(4;14) Positive Patients.... 126

Figure C2.3 – The Influence of t(4;14) and *FGFR3* Expression on Overall Survival.... 129

Figure C2.4 – Description of der(14) Screening Assays 132

Figure C2.5 – The Poor Outcome Associated with t(4;14) is Independent of FGFR3 Expression and Breakpoint Type 137

Figure C2.6 – Genes Flanking the t(4;14) Breakpoints on 4p16 140

Figure C2.7 – RE-IIBP Expression in BMMC 147

Figure C3.1 – MMSET Splicing and Protein Variants 158

Figure C3.2 – Localization of N-Terminally Tagged Wild-type MMSET Variants 160

Figure C3.3 – Localization of C-terminally Tagged MMSET Protein Variants 162

Figure C3.4 – Truncated Protein Products are Produced from the Myeloma Specific MB4-2 and MB4-3 Transcripts 164

Figure C3.5 – Alternative Translation Sites Produce Mis-Localized MMSET Variants 166

Figure C3.6 – The N-terminus of MMSET regulates its Localization Pattern. 169

Figure C3.7 – FRAP Recovery Curves of Wild-type and Breakpoint Variants. 172

Figure C3.8 – Domain Architecture of NSD Paralogs and Orthologs 175

Figure C3.9 – The Evolution of the NSD Family 177

Figure C3.10 – Evolution of the SET2 Histone Methyltransferase Family 179

Figure C3.11 – N-Terminal Conservation between WHSC1/MMSET and WHSC1L1 182

Figure C3.12 – Multiple Sequence Alignment of WHSC1/MMSET Orthologs 185

Figure C3.13 – Multiple Sequence Alignment of the C-Terminus of Human NSD Paralogs 189

Figure C4.1 – The Breakpoint Cloning Strategy 200

Figure C4.2 – The Different IgH Translocation Models in Myeloma 207

Figure C4.3 – Breakpoint Diagrams 210

Figure C4.4 – Characterized t(4;14) Breakpoints are not Randomly Distributed 220

Figure D.1 – Models of t(4;14) Myelomagenesis 261

List of Symbols, Nomenclature, or Abbreviations

ACF1	ATP-dependent chromatin assembly factor
ACH	Acondroplasia
ALL	Acute Lymphocytic Leukemia
AML	Acute Myelogenous Leukemia
AREG	Amphiregulin (schwannoma-derived growth factor)
ASH1L	ash1 (absent, small, or homeotic)-like (Drosophila)
ATRA	All-Trans Retinoic Acid
AWS	Always With SET domain
B23	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)
B2-M	Beta-2-Microglobulin
BAC	Bacterial Artificial Chromosome
BCNU	Bischloroethyl nitrosourea
BCR-ABL	Breakpoint Cluster Region protein- Abelson murine Leukemia viral oncogene homolog 1 fusion
BLAST	Basic Local Alignment Search Tool
BM	Bone Marrow
BMMC	Bone Marrow Mononuclear Cells
bp	Base Pair
BROMO	Bromodomain
C-	Carboxy
C+AN	Crouzon with acanthosis nigricans

CC	Coiled-Coil
CCND1	Cyclin D1
CCND3	Cyclin D3
cDNA	Complimentary DNA
CDR2	Complement Determining Region 2
CDR3	Complement Determining Region 3
CGH	Comparative Genomic Hybridization
CH	Constant region (Immunoglobulin Heavy Chain)
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
cIg-FISH	Cytoplasmic Immunoglobulin enhanced FISH
CI	Confidence Interval
CK2	Casein Kinase 2
CML	Chronic Myeloid Leukemia
CMV	Cytomegalovirus
CRS	Cranisynostoses associated or without other limb malformations
CSDA	Cold Shock Domain protein A
CSR	Class Switch Recombination
Ct	Cycle threshold
D-TACC	Drosophila melanogaster TACC
DJ	Diversity, Joining
DNA	Deoxyribonucleic Acid
DNMT3a	DNA (cytosine-5-)-Methyltransferase 3 alpha

DNMT3b	DNA (cytosine-5-)-methyltransferase 3 beta
ECOG	The Eastern Cooperative Oncology Group
Edm	Edmonton
EF-1α	Elongation Factor 1-alpha
EMSA	Electrophoretic Mobility Shift Assay
EST	Expressed Sequence Tag
FAB	French, American, British
FGFR1	Fibroblast Growth Factor Receptor 1
FGFR2	Fibroblast Growth Factor Receptor 2
FGFR3	Fibroblast Growth Factor Receptor 3
FGFR4	Fibroblast Growth Factor Receptor 4
FISH	Fluorescent In Situ Hybridization
FOG-1	Friend of GATA-1
FRAP	Fluorescence Recovery After Photobleaching
FYVE	Fab-1, YGL023, Vps27, and EEA1 domain
G-banding	Giemsa chromosome banding
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GATA-1	Globin Transcription factor 1
GEP	Gene Expression Profiling
GFP	Green Fluorescent Protein
GST	glutathione-S-transferase

H3-K4	Histone 3-Lysine 4
H3-K9	Histone 3-Lysine 9
H3-K36	Histone 3-Lysine 36
H4-K20	Histone 4-Lysine 20
HAT	Histone Acetyltransferase
HCH	Hypochondroplasia
HDGF	Hepatoma-Derived Growth Factor
HDT	High Dose Therapy
HGMD	Human Gene Mutation Database
HIV	Human Immunodeficiency Virus
HMG	High Mobility Group
HMT	Histone Methyltransferase
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HR	Hazard Ratio
HRAS	v-Ha-ras Harvey Rat Sarcoma viral oncogene homolog
HYPB	Huntingtin interacting Protein B
Iμ	Intervening region mu
IFM	Intergroup Francophone du Myelome
IgH	Immunoglobulin Heavy Chain
IgL	Immunoglobulin Light Chain
IL-6	Interleukin 6
IL6R	Interleukin 6 Receptor
IMID	Immunomodulatory Drugs

INF-α2	Interferon alpha 2
ING2	Inhibitor of Growth family, member 2
IRF4	Interferon Regulatory Factor 4
JH	Joining Segment (Immunoglobulin Heavy Chain)
kb	Kilobase
kDa	Kilo Dalton
KLF4	Kruppel-Like Factor 4
KRAS	v-Ki-ras2 Kirsten Rat Sarcoma viral oncogene homolog
LDH	Lactate Dehydrogenase
LETM1	Leucine Zipper-, EF-hand-containing Transmembrane Protein 1
LSI	Locus Specific Identifier
LSM	Laser Scanning Microscope
M-Proteins	Monoclonal proteins
MAEA	Macrophage Erythroblast Attacher
MAF	v-maf Musculoaponeurotic Fibrosarcoma oncogene homolog
MAFB	v-maf Musculoaponeurotic Fibrosarcoma oncogene homolog B
MAL	mal, T-cell differentiation protein
MAPK	Mitogen-Activated Protein Kinase
Mb	Megabase
MB	Mobilized Blood
MB4-1	Major Breakpoint chromosome 4 region 1

MB4-2	Major Breakpoint chromosome 4 region 2
MB4-3	Major Breakpoint chromosome 4 region 3
MB4-4	Major Breakpoint chromosome 4 region 4
ME	MMSET Exon
MEK	Mitogen-activated protein kinase kinase
Mes-4	Maternal Effect Sterile 4
MGUS	Monoclonal Gammopathy of Undetermined Significance
MI	MMSET Intron
MM	Multiple Myeloma
MMSET	Multiple Myeloma SET domain containing protein
MMSET I	MMSET type I
MMSET II	MMSET type II
MMSET III	MMSET type III
mRNA	messenger Ribonucleic Acid
MSH6	mutS homolog 6 (E. coli)
MSI-H	Microsatellite Instability High
MSI-L	Microsatellite Instability Low
Mw	Molecular Weight
N-	Amino
NCBI	National Center for Biotechnology Information
NELF	Negative Elongation Factor
NELF-A	Negative Elongation Factor-Subunit A
NGFRAP1	Nerve Growth Factor Receptor (TNFRSF16) Associated Protein 1

NR	Nuclear Receptor
NSD1	NR binding SET Domain containing protein 1
NSD2	NR binding SET Domain containing protein 2
NSD3	NR binding SET Domain containing protein 3
NRAS	Neuroblastoma Rat Sarcoma viral (v-ras) oncogene homolog
nt	Nucleotide
ORF	Open Reading Frame
PAC	P1-dervived Artificial Chromosome
PB	Peripheral Blood
PBMC	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PBSC	Peripheral Blood Stem Cell
PCL	Plasma Cell Leukemia
PCR	Polymerase Chain Reaction
PHD	Plant Homeodomain
pI	Isoelectric Point
PKC	Protein Kinase C
PLSD-SD	Platyspondylic Lethal Skeletal Dysplasia, San Diego type
PML-RARα	Promyelocytic Leukemia protein- Retinoic Acid Receptor, alpha fusion
POLN	Polymerase (DNA directed) nu
PtdIns(3)P	Phosphatidyl Inositol-3-Phosphate
PtdIns(5)P	Phosphatidyl Inositol-5-Phosphate

PTEN	Phosphatase and Tensin homolog
PWWP	Proline-Trypophan-Trypophan-Proline
qRT-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
RAF	v-raf-1 murine leukemia viral oncogene homolog 1
RAS	Rat Sarcoma
RASGRP1	RAS guanyl releasing protein 1
RBMS1	RNA Binding Motif, Single stranded interacting protein 1
RE-IIBP	Response Element II Binding Protein
REL	Relative Expression Level
RFI	Relative Fluorescence Intensity
RING	Really Interesting New Gene
RIZ	Retinoblastoma Interacting Zinc finger protein
RNA	Ribonucleic Acid
ROI	Region of Interest
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
Sα	Switch region alpha
Sγ	Switch region gamma
Sμ	Switch region mu
SADDAN	Severe Achondroplasia with Developmental Delay and Acanthosis Nigricans
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SET	Su(var)3-9, Enhancer of zeste, Trithorax
sIg	Surface Immunoglobulin
siRNA	Short Interfering RNA
SKY	Spectral Karyotyping
SLBP	Stem Loop Binding Protein
SMART	Simple Modular Architectural Research Tool
sMM	Smoldering Multiple Myeloma
SPN	Sialophorin (CD43)
SPON2	Spondin 2, extracellular matrix protein
SUV39	Suppressor of variegation 3-9
TACC1	Transforming Acidic Coiled-Coil gene 1
TACC2	Transforming Acidic Coiled-Coil gene 2
TACC3	Transforming Acidic Coiled-Coil gene 3
Tat	Transcriptional Activator (Human Immunodeficiency Virus)
TDI	Thanatophoric Dysplasia type I
TDII	Thanatophoric Dysplasia type II
TP53	Tumor Protein p53 (Li-Fraumeni syndrome)
UTR	Untranslated Region
UPMGA	Unweighted Pair Group Method with Arithmetic Mean
UV	Ultraviolet
VAD	Vincristine, Adriamycin, Dexamethasone

VBMCP	Vincristine, BCNU, Melphalan, Cyclophosphamide, Prednisone
VDJ	Variable, Diversity, and Joining
VH	Variable segment (Immunoglobulin Heavy Chain)
VJ	Variable and Joining
WHS	Wolf-Hirschhorn Syndrome
WHSC1	Wolf-Hirschhorn Syndrome Candidate gene 1
WHSC1L1	Wolf-Hirschhorn Syndrome Candidate gene 1, Like 1
WHSC2	Wolf-Hirschhorn Syndrome Candidate gene 2
WM	Waldenström's Macroglobulinemia
YAC	Yeast Artificial Chromosome
ZFP36L1	Zinc Finger Protein 36, C3H type-Like 1

Introduction

I.1.1 – The Human Face of Multiple Myeloma

The clinical entity called multiple myeloma is characterized by the accumulation of plasma cells, the antibody producing cells, in the bone marrow. This increase in the percentage of plasma cells in the bone marrow from the normal range of 2% to 10-100% is associated with several clinical features. First, in most patients a monoclonal protein spike is present in the serum as the malignant plasma cells produce large quantities of a single antibody. Second, the replacement of normal bone marrow cells with plasma cells often impairs normal hematopoiesis. This can result in anemia as red blood cell production is impaired and decreased immune function as white blood cell production is also impaired. Third, osteolytic lesions are commonly seen in the skull, ribs, vertebrae, and femur. Currently, myeloma is incurable and deaths typically occur as a result of renal failure, a pathological effect of the monoclonal protein, or infections as the immune system is compromised by both the cancer and treatment.

Dear Dr. X,

I am referring Mrs. Z to you for further investigation of what I believe to be multiple myeloma. Mrs. Z is a 58 year old retired school teacher with no previous health conditions. She has been a regular patient in my family practice clinic for the majority of my 15 year career. She recently visited my clinic complaining of severe back pain and consistent fatigue. We performed a general work-up on February 11, which revealed; Hemoglobin 93, Calcium 2.98, Albumin 37, Creatinine 220, Total serum protein 100 g/L, and a monoclonal protein spike of 54 g/L. I called and requested that she return to the clinic the following week. Today we had X-rays taken of the skull and chest, which revealed numerous lytic lesions. Furthermore, I note that she has lost 2.5 cm in height since the summer of 2003. Based on these criteria, I discussed with Mrs. Z my intention to refer her to the regional cancer care centre and my belief that she may have multiple myeloma. Mrs. Z. was very distraught and upset with this news. She recently lost her husband to colon cancer and was extremely worried about the prognosis associated with myeloma. I called her this afternoon and informed her that we arranged an appointment with you a week from today.

Sincerely,

Dr.Y
February 15, 2005

Dear Dr. Y

Mrs. Z and her youngest son were seen in our clinic on February 22. They were very worried about the potential diagnosis of myeloma and had obviously spent time on the internet researching the disease and treatment options. As you suspected this lady has multiple myeloma. Our investigations are detailed below:

Pathology: Peripheral blood – Rouleux and rare plasma cells are seen
Bone Marrow – Total cellularity is increased and an infiltration of plasma cells representing 65% of the marrow cellularity is present
Laboratory: Hemoglobin 92, Calcium 2.96, Albumin 37, Creatinine 218, Beta-2-Microglobulin 4.8, Total protein 99 g/L, M-protein 57 g/L
Special Tests: Cytogenetics – Karyotype analysis revealed a hypo-diploid karyotype
42, XX, del(1)(p12p32), del(6)(q16), -11, -11, -13, -14.
Interphase FISH confirmed the del(13), del(14), and identified a t(4;14)(p16;q32) translocation.
Labeling Index – 1.3%
MRI/X-ray – Extensive lytic disease is noted with multiple punched out lesions in the skull and long bones. A complete compression of L3 and partial of L4 is noted.

Diagnosis – This lady has IgA-lambda multiple myeloma (Durie-Salmon stage IIIB) with several poor prognosis markers including Beta-2-microglobulin >3.0, deletion 13, and t(4;14).

I had the opportunity to sit down with Mrs. Z and her son after the initial consultation and discussed the diagnosis of myeloma and the potential treatment opportunities. As expected they were very upset. With her level of bone disease I suggested we start her on Pamidronate immediately to deal with her bone disease and I've referred her to our orthopedic surgeons to determine if she would by a candidate for vertebral stabilization. Furthermore, we provided her with information regarding the treatment options and rescheduled an appointment for today at which time all of the investigation results would be available. Mrs. Z returned today with her son and we discussed potential treatment options in respect to the poor prognostic features of her myeloma. Typically for a patient in good health, like Mrs. Z, we would recommend an autologous transplant. However, given her poor prognostic markers, Mrs. Z has elected to enroll in one of our clinical trials, specifically Velcade with or without Dexamethasone as front line therapy. Velcade has shown promising utility in relapsed patients. If she does not respond or relapses with Velcade resistance she would still be eligible for our Revlimid trial for relapsed refractory patients without previous Thalidomide or IMID therapy. I am happy to note that Mrs. Z was excited to have access to these new drugs and was looking forward to initiating therapy. Her positive attitude may prove to be more beneficial then any therapy we can provide. She will meet with our research nurse later this afternoon and should receive her first cycle of therapy tomorrow.

Sincerely,

Dr. X
February 28, 2005

The short clinical extraction noted above is reflective of many clinical records that I've read over the course of my degree. Unfortunately, many patients start with the optimism of Mrs. Z but due to the highly aggressive form of multiple myeloma conferred by t(4;14)(p16;q32) they often succumb to the disease quite quickly. This unfortunate truth reflects the importance of much of the work listed in this thesis. It is only through an understanding of the biological mechanisms that we are likely to develop functional and specific therapeutics for patients like Mrs. Z.

I.2 – General Background on Multiple Myeloma

I.2.1 – Demographics in Canada

Multiple myeloma is a relatively rare disease with a yearly incidence rate of 6 per 100 000 in Canada[i]. It represents 1.2% of all recorded cancers and 14.3% of all hematological malignancies. However, it accounts for 19.0% of the deaths due to hematological cancers. This malignancy remains one of the worst cancers with a death to case ratio of 0.67. The only other common cancers with worse or comparable death to case ratios in Canada are esophagus, pancreas, lung, brain, stomach, and ovary[ii]. Multiple myeloma is a disease of the elderly, with a median age at diagnosis of 67 years[1]. Furthermore, the incidence is higher in males (53.9% of cases) than females (46.1% of cases). Interestingly, this male/female incidence rate reported in the most recent Canadian cancer statistics is not reflective of previous numbers. Typically the incidence rate follows a 1/3 female 2/3 male distribution, and this distribution is found in our original t(4;14) cohort of 208 myeloma patients collected between 1994 and 2002. It may be warranted to determine if this ratio is changing as it may reflect behavioral and environmental changes associated with multiple myeloma.

I.2.2 – Differential Diagnosis and Follow-up

Multiple myeloma is a member of a range of clinical disorders characterized by a γ-globulin spike on serum electrophoresis. These monoclonal gammopathies and related disorders include monoclonal gammopathy of undetermined significance (MGUS), multiple myeloma, plasmacytomas, plasma cell leukemia, Waldenström's macroglobulinemia, and primary amyloidosis. The differential diagnosis of these related disorders should be aided by the simplified set of diagnostic criteria recently released by The International Myeloma Working Group summarized in Table I.1 and the commonly accepted diagnostic criteria for Waldenström's Macroglobulinemia, and primary amyloidosis[2,3]. The diagnostic criteria are quite stringent; however, inevitable confusion often results as patients may, over time, transition from one diagnosis to another. Furthermore, some patients may need to be monitored for a period of time before a diagnosis is confirmed. This is common when trying to differentiate Waldenström's macroglobulinemia from IgM MGUS or determining if a patient has asymptomatic myeloma, MGUS, or symptomatic myeloma.

---

[i] Numbers are from Appendix I of Canadian Cancer Statistics 2004 (www.cancer.ca) and reflects actual data collected across Canada for the year 2000, published online by the National Cancer Institute of Canada, the Canadian Cancer Society, Statistics Canada, and Health Canada.

[ii] Listed in decreasing order

The majority of myeloma patients present with symptomatic myeloma, however, many transition to myeloma after a previous diagnosis of plasmacytoma or MGUS. Approximately 50% of patients initially diagnosed with a solitary plasmacytoma of the bone will transition to myeloma[2]. Typically this occurs in 3 to 4 years but may take up to 15 years occur. Furthermore, patients with recurrent plasmacytomas or plasmacytomas occurring in multiple sites have a high risk of transformation. Interestingly, only 16% of patients with extramedullary plasmacytomas will transform to myeloma[4]. Patients with an initial diagnosis of MGUS may transition to Waldenström's macroglobulinemia (if IgM isotype), lymphoma (also commonly IgM isotype), asymptomatic myeloma, symptomatic myeloma, plasmacytoma or amyloidosis[5,6]. Overall 1% of MGUS patients will transition per year and ultimately 25% will transition with a median time to progression of 10 years. The relative risk of a MGUS patient transitioning is 7.3.

Specifically the relative risk of developing Waldenström's macroglobulinemia, myeloma, amyloidosis, or plasmacytomas is 46, 25, 8.4, and 8.5, respectively[6]. The incidence of transition from asymptomatic to symptomatic myeloma is not well defined; however, these patients require regular monitoring. Regular monitoring of the serum M-protein and periodic assessment of the level of bone marrow plasmacytosis is necessary to determine the transition from MGUS to asymptomatic or symptomatic myeloma and to confirm a specific diagnosis.

Table I.1 – Differential Diagnosis of Monoclonal Gammopathies

Monoclonal Gammopathy of Undetermined Significance (MGUS)

- Serum M-protein <30 g/l
- Bone marrow plasma cells <10%
- No evidence of alternative B-cell disorders
- No related organ or tissue impairment (CRAB criteria)
    - **C**alcium increased → serum Calcium >2.75 mmol/l
    - **R**enal insufficiency → serum Creatinine >173 mmol/l
    - **A**nemia → Hemoglobin <100 g/l
    - **B**one Lesions → Lytic lesions or compression fractures

      *plus hyperviscosity, amyloidosis, and more than two recurrent bacterial infections in 1 year.

Asymptomatic Myeloma (formerly smoldering myeloma)

- Serum M-protein ≥30 g/l
- With **or** without bone marrow plasma cells ≥10%
- No related organ or tissue impairment (CRAB criteria)

Symptomatic Myeloma – Multiple Myeloma

- M-protein in serum or urine

  *97% of patients have a detectable M-component at diagnosis.

  *Typically >30 g/l but may be lower in 40% of patients at diagnosis
- Clonal bone marrow plasma cells (typically ≥10%, range 5-100%)
- Presence of related organ or tissue impairment (CRAB criteria)
    - **C**alcium increased → 20%
    - **R**enal insufficiency → 20%
    - **A**nemia → 70%
    - **B**one Lesions → 80%, by X-ray

Non-Secretory Myeloma

- No M-protein in serum or urine
- Bone marrow clonal plasma cells >10% or plasmacytomas
- Presence of related organ or tissue impairment (CRAB criteria)

Plasmacytoma

- Histological evidence of a monoclonal plasma cell tumour
- No evidence of multiple myeloma

- Three common types exist → Solitary Plasmacytoma of the Bone
  → Extramedullary Plasmacytoma
  → Multiple Solitary Plasmacytomas (± Recurrent)

Plasma Cell Leukemia

- Absolute peripheral blood plasma cell count >2.0 x 10^9/l
- Plasma cells representing >20% of the manual differential count
  - Primary Plasma Cell Leukemia → Identified at diagnosis
  - Secondary Plasma Cell Leukemia → Transition from diagnosis of multiple myeloma

Waldenström's Macroglobulinemia – Lymphoplasmacytic lymphoma†

- Presence of an IgM M-protein spike (variable levels)
- Bone marrow infiltration with lymphoplasmacytic lymphoma cells with an intertrabecular distribution and $\text{sIg}^{+}\text{CD19}^{+}\text{CD20}^{+}\text{CD22}^{+}\text{CD79}^{+}$ phenotype

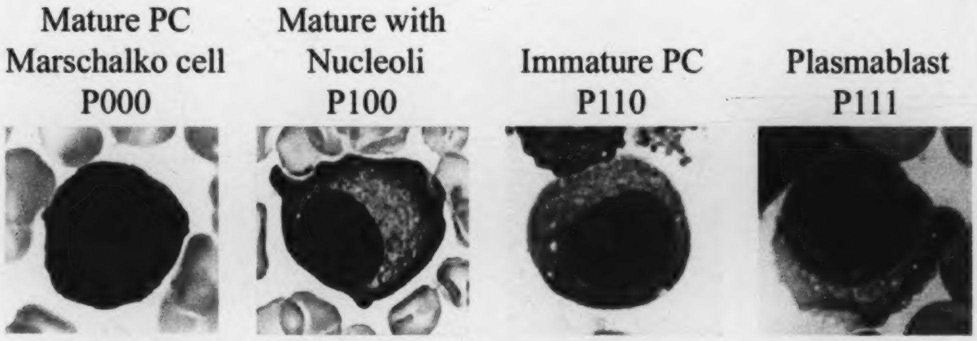
Primary Amyloidosis - Amyloidosis†

- Amyloid side effects
  - Nephrotic
  - Congestive heart failure
  - Carpal tunnel syndrome
  - Neuropathy
  - Autonomic
- Presence of serum and urine light chains
- Congo Red positive bone marrow, subcutaneous fat, kidney, or liver biopsies
- Bone marrow plasma cells <10%

† Diagnostic criteria for Waldenström's Macroglobulinemia and Primary Amyloidosis presented by Gertz et al. at the 2004 Annual Meeting of the American Society of Hematology[3].

I.2.3 – Phenotype and Hierarchy of Myeloma Cells

One of the principle diagnostic criteria for multiple myeloma is the bone marrow infiltration of morphological plasma cells. The typical/mature plasma cell or Marschalko cell is a large terminally differentiated B-cell characterized by the absence of nucleoli, a low nuclear/cytoplasm ratio (<0.6), and condensed chromatin. On average this cell type represents at least 50% of the plasma cells in the bone marrow of myeloma patients[7]. However, the bone marrow of most patients will contain plasma cells of other differentiation stages. The terminology varies between groups, however, the typically defined plasma cell subsets in order of differentiation are plasmablast, immature, intermediate, and mature. In general, the presence of nucleoli and fine chromatin differentiate the plasmablast and immature plasma cell from the typical mature plasma cell. However, in the study of Goasguen et al. a subset of mature plasma cells with a nucleolus were relatively common and aided in determining patient prognosis[7]. In this study, plasma cells fitting the description of a plasmablast, immature plasma cell, and mature plasma cells with or without nucleoli represented 93% of the scored plasma cells (Figure I.1). Patients with mature plasma cells without nucleoli representing more than 66% of the plasma cells had a good prognosis. Those with less than 66% could be split into two groups with progressively worse prognosis. Patients having more mature plasma cells with nucleoli than plasmablasts and immature plasma cells had an intermediate prognosis, while those with a high frequency of plasmablasts and immature plasma cells had the worst prognosis. This observation was expected as the poor prognosis of patients with more than 2% plasmablasts is well established[8,9]. The presence of plasma cells with nucleoli is often a defining characteristic of multiple myeloma, though the percentage of cells with nucleoli and their differentiation stage varies widely between patients.


Mature PC
Marschalko cell
P000

Mature with
Nucleoli
P100

Immature PC
P110

Plasmablast
P111

Figure I.1 – Typical Plasma Cell Morphologies Observed in Myeloma Patients

Images were kindly provided by Jean Goasguen (Université de Rennes, Rennes, France). The typical plasma cell morphologies characterized by Goesguen et al. are shown[7]. Their sub-grouping algorithm characterizes plasma cells based on three criteria. The first sub-grouping, P1 or P0, is based on the presence or absence of nucleoli. The second sub-grouping, Px1 or Px0, is based on the presence or absence of blastic chromatin (fine uncondensed chromatin). The third sub-grouping, Pxx1 or Pxx0, reflects a nuclear/cytoplasmic ratio >0.6 or <0.6, respectively. Though this algorithm identifies eight different sub-groups only the P000, P100, P110, and P111 sub-groups are shown as these represent approximately 93% of the plasma cells characterized in myeloma patients using this method. Typically, the P000 sub-type represents >50% of the bone marrow plasma cells.

Myeloma plasma cells can also be characterized based on their immunophenotype. The distinction between myeloma and normal plasma cells can be made on several features assessed by flow cytometry[10-13]. In general, normal plasma cells are CD19^{+}, CD38^{+++}, CD45^{+}, CD56$^{\pm}$, CD126^{-}, and CD138^{+} with normal DNA content. Most myeloma plasma cells will be CD19^{-}, CD38^{++}, CD45$^{\pm}$, CD56^{++}, CD126^{+}, CD138^{++} with an aneuploid DNA content. Other markers like CD20, CD28, CD33, CD117, and sIg may help but they are only informative in 6-30% of patients[12]. In general plasma cells of any origin can be identified by the co-expression of CD38 and CD138. Subsequently, normal or myeloma plasma cells can be identified within this subset of CD38/CD138 co-expressing cells based on the expression of CD19 and CD56[10]. Most normal plasma cells will be CD19^{+}/CD56^{-} while myeloma plasma cells will be CD19^{-} /CD56^{+}. Importantly, co-expression of CD19 and CD56 is only found in 4% of myeloma patients. The expression of CD45 is often heterogeneous within plasma cell populations of individual patients. However, this may reflect distinct populations as most proliferative cells within the plasma cell compartment express high levels of CD45[14]. Therefore based on immunophenotype it is possible to identify plasma cells, differentiate normal and myeloma plasma cells, and to identify proliferative subsets.

Although immunophenotyping can differentiate light chain restricted plasma cells, this does not conclusively prove clonality. The only way to unequivocally differentiate the monoclonal myeloma plasma cells from normal polyclonal plasma cells is by molecular assessment of the immunoglobulin heavy and light chain rearrangements. The genetic rearrangement and subsequent somatic hypermutation of the IgH VDJ or IgL (kappa or lambda) VJ segments provides each myeloma patient with a unique molecular signature[15,16]. Using PCR based techniques to detect this molecular signature several groups have identified clonal cells in the bone marrow and peripheral blood with different immunophenotypes. In particular, CD19 positive clonal cells exist in the peripheral blood, however, the frequency of these cells is of substantial debate[17-24]. Several groups have suggested that these cells represent drug resistant myeloma stem cells that recapitulate the bone marrow plasma cells after therapeutic eradication. However, the presence of this unique molecular signature does not mean the clonal plasma cell or B-cell is transformed, but simply identifies cells originating from a common progenitor.

I.3 - The Underlying Genetics of Multiple Myeloma

A number of the common genetic events associated with a malignant phenotype are found in myeloma. These include activating or inactivating mutations and deletions of several common oncogenes and tumour suppressors like *TP53*, *RAS*, *PTEN*, and others. Moreover, duplications and deletions of chromosomal segments are commonly observed. These events may promote a malignant phenotype by altering the gene dosage of proto-oncogenes or tumour suppressors. Alternatively, complete gains and losses of chromosomes, which are common in myeloma, may mediate a similar gene dosage effect though on a much larger scale. Finally, a variety of structural aberrations occur, with translocations involving chromosome 1 or chromosome 14 being the most common. Which lesions are primary and secondary events are slowly being elucidated. It is assumed that primary events will be detected in the pre-malignant MGUS stage and/or the majority of the plasma cells at diagnosis. Alternatively, lesions detected in a subset of plasma cells at diagnosis or those detected at relapse are believed to be secondary events.

I.3.1 – Common Oncogenes and Tumour Suppressors in Myeloma

The most common genetic change observed in human cancers is the loss of p53 function as a result of mutations and deletions of *TP53*[25]. In myeloma the incidence of *TP53* mutations and deletions is not well defined. Since; the majority of studies only looked at deletions or mutations, and rarely was the incidence of both events determined. Inactivation of p53 can result from a variety of mechanisms though deletions and inactivating mutations are the most common. Several groups have investigated the incidence of gross deletions of the *TP53* locus in myeloma cells using fluorescent in situ hybridization (FISH). Two large studies with untreated patients observed similar frequencies of *TP53* deletions and when combined deletions are detected in 47/450 (10.4%) of myeloma patients [26,27]. In the study by Fonseca et al. deletions of the *TP53* locus were associated with t(14;16)(q32;q23) and though statistically insignificant, an association with t(4;14) was suggested[26]. Though this suggested association with t(4;14) was not observed by Chang et al., this could be due to their poorly designed t(4;14) FISH assay which only detects a single derivative chromosome[27]. Though not noted by Fonseca et al., in most studies the deletions of *TP53* are mono-allelic suggesting that a wild-type allele may still exist[27,28]. Furthermore, deletions of *TP53* appear to be progression, not initiation, events as the percentage of plasma cells with deletions varies, being below 50% in 30% of patients, and the incidence of deletions is higher at relapse[26,28,29]. Similarly, mutations of *TP53* are more common in advanced disease stages and in some patients are detectable at relapse but not diagnosis[30]. Though mono-allelic loss of *TP53* should not ablate p53 function this event is a poor prognostic indicator in multiple myeloma[26-28,31]. Therefore, it appears as if haplo-insufficiency of p53 in myeloma is a poor prognostic indicator.

Activating mutations of RAS proto-oncogenes are one of the most common characterized mutations in human cancers. Similarly, activating mutations of *NRAS* or *KRAS* but not *HRAS* are relatively common in newly diagnosed cases of myeloma (98/251 or 39%)[32-34]. The frequency of the various mutations at diagnosis is independent of disease stage as equal frequencies are seen in all Durie-Salmon stages[32,34]. Moreover, ras mutations are only found in approximately 5% of MGUS patients, suggesting they promote the development of the malignant myeloma phenotype [32]. Though this suggests ras mutations may be initiating events, not all patients with mutations at relapse had detectable mutations at diagnosis[33,34]. The increased prevalence of ras mutations at relapse may reflect clonal selection. In two studies, multiple mutations were detected in some patients (Bezieau et al.; Mean 1.6, Range 1-4)[33,34]. However, some of the mutations appear to reside in rare sub-clones as they are not always evident by conventional direct sequencing[33]. Moreover, one case with multiple mutations in the bone marrow developed a plasmacytoma with only one of the two mutations[34]. Similarly, mutations were only detected in the extramedullary tumours of 3 patients with intramedullary and extramedullary plasma cell tumours[32]. In the case of t(4;14) positive myeloma cell lines an inverse relationship exits between the presence of activating mutations of *FGFR3* and *NRAS* or *KRAS*[35]. However, in primary myeloma samples this relationship does not appear to exist; as 5 of 5 tested patient samples did not have mutations of either gene[36]. Furthermore, no mutations of *NRAS* or *KRAS* were observed in 17 untreated t(4;14) positive samples[32]. Therefore, mutations of *FGFR3*, *NRAS*, or *KRAS* are likely required for the establishment of a t(4;14) positive cell line but are not likely required for growth in patients.

One of the most commonly mutated tumour suppressors in human malignancies is *PTEN*[37]. In the case of multiple myeloma very little research exists regarding the involvement of this tumour suppressor. To date only nine of the common myeloma cell lines have been screened for *PTEN* mutations or deletions[iii][38-40]. Of these nine cell lines, two have bi-allelic deletions and one a mono-allelic deletion. Interestingly, two (OPM-2 and JIM3) of four (KMS-11 and NCI-H929) tested t(4;14) positive cell lines have *PTEN* mutations. The cell lines without wild-type PTEN, OPM-2 and Delta47, are sensitive to drugs that inhibit the PTEN associated pathway[40]. Moreover, OPM-2 cells expressing high levels of PTEN lose their ability to form tumours in SCID mice[41]. Further investigation of *PTEN* in primary myeloma samples is required to elucidate its involvement in myelomagenesis. In particular, it may be pertinent to determine the occurrence of PTEN inactivation in t(4;14) positive samples since 50% of t(4;14) positive cell lines have *PTEN* mutations.

[iii] The known EBV positive HS-Sultan and ARH-77 cell lines were excluded as are several of the rare cell lines like UCLA #1, UCLA #2, AF-10 where EBV status and functional characterization are lacking.

I.3.2 – Ploidy and Myeloma

The sub-grouping of myeloma patients based on ploidy characteristics is now widely accepted. In 1984, Lewis and McKenzie correctly identified the majority of what are now the widely accepted karyotypical abnormalities in myeloma[42]. In a meta-analysis of 27 karyotypes they concluded that structural changes of chromosome 1 and 14, trisomies of chromosomes 3, 5, 7, 9, 11 and monosomies of chromosome 8 and 13 were the most common cytogenetic abnormalities in myeloma. Though refinements with improved techniques and larger cohorts have occurred, the observations of Lewis and McKenzie still hold true today. Several different techniques with distinct advantages and disadvantages are used to determine ploidy/DNA content; including conventional G-banding or spectral karyotyping, comparative genomic hybridization (CGH), flow cytometry, and interphase FISH.

The most commonly used technique to determine ploidy status is karyotype analysis by either conventional G-banding or fluorescence-based methods such as spectral karyotyping (SKY). These techniques can identify chromosomal gains, losses, structural abnormalities and large internal duplications and deletions. The major drawback of either method of karyotype analysis is the need for metaphase spreads.

Unfortunately, very few myeloma plasma cells are proliferative *in vitro* and thus, the metaphase spreads required for karyotypic analysis are only generated in approximately 30% of patients[43]. Therefore, the results generated by karyotype analysis may not be representative of the entire myeloma population. As suggested by Lewis and McKenzie several common cytogenetics changes are seen in myeloma patients. These include trisomies of chromosomes 3, 7, 9, 11, 15, 19, 21 and monosomies of chromosomes 8, 13, 14, 16, 17, 22, X[43-46]. Structural changes are most frequent in chromosome 1 though changes at 6q, 7p, 8q, 11q, and 14q are commonly observed. By karyotype analysis patients can be sub-grouped into hypodiploid, pseudodiploid, hyperdiploid and near tetraploid. These groups predict for a differential clinical outcome with a hypodiploid karyotype predicting a poor prognosis[43,44]. The most extensively investigated cytogenetics finding is chromosome 13 deletion, which is associated with poor prognosis and an increased risk of MGUS to myeloma transition[47-57]. Ultimately, conventional cytogenetics is capable of identifying prognostic subgroups, but the inability to generate metaphase spreads in some patients and the limited resolution limit its wide spread utility.

The technique of CGH is one of the best ways to determine DNA content and specific changes. In this assay the tumour DNA is competitively hybridized to target DNA to determine specific gains and losses. This technique can be used on all patients, however, a tumour DNA content of 50% is needed for accurate results. In the myeloma situation this should not be an obstacle given the wide spread use of magnetic purification systems to purify plasma cells from bone marrow mononuclear cells (BMMC). Furthermore, the resolution of CGH is significantly better than conventional cytogenetics so small amplifications and deletions, undetectable by conventional cytogenetics, are detected. To date only conventional CGH has been used in myeloma and similar to the flow cytometry results 60-70% of patients have chromosomal changes[58-60]. In general, the gained and lost chromosomes are similar to those identified by conventional cytogenetics with gains of 3, 5, 7, 9, 11, 15, 19 and losses of 13, and X commonly identified. However, CGH provides more specific information regarding the regions of gain or loss. In particular, gains of 9, 11, and 15 were resolved to 9q31qter, 11q23qter, and 15q23qter and additional minimal regions of gain at 1q21q23, 1q25q31, and 6p21were identified[59,60]. Common regions of deletion were defined at 1p21, 6q21, 8p21, and 13q14q21[59,60]. Moreover, the detection of DNA loss by CGH, most likely hypodiploid cases, was an indicator of poor prognosis[58]. Further resolution will be possible with the use of array-CGH, which has substantially higher resolving power. Unfortunately, CGH results are only reflective of the average DNA content within a sample. So differences between sub-clones will not be observed. However, in regards to resolution and applicability, CGH is the best method for assessing global DNA content and identifying specific regions of interest.

The DNA content of myeloma cells can be determined by multi-colour flow cytometry. This method involves the immunphenotypic identification of myeloma plasma cells (CD38$^+$/CD138$^+$/CD19$^-$/CD56$^+$) followed by DNA content assessment of the cells within this population. Using this method ~60% of myeloma cases have detectable DNA content differences[61,62]. Similar to the observations using conventional cytogenetics the non-hyperdiploid cases have a poor prognosis[61]. One major advantage of this technique is that it can be performed on all patients. However, this technique, like CGH, only gives a global view so small changes (+/-1-3 chromosomes) between subclones may not be detected, but unlike CGH, hypodiploid and hyperdiploid subclones can be detected within the same sample.

Interphase FISH can also be used to determine ploidy. Although, this technique can be used on all patients, only a limited number of loci can be investigated in each cell due to fluorescence detection limits. Therefore, interphase FISH is a good technique for measuring specific changes in all cells, but since only a limited number of loci are investigated the risk of misclassification is high.

I.3.3 – Translocations in Myeloma

Two major types of translocations exist in multiple myeloma. The most common are the immunoglobulin translocations which involve the IgH, IgL-kappa, and IgL-lambda loci located at 14q32, 2p11, and 22q11, respectively. Alternatively, a plethora of non-immunoglobulin translocations occur involving a variety of chromosomes. The incidence of IgH translocations has been extensively studied with conventional cytogenetics, metaphase FISH, interphase FISH, and southern blotting. Surprisingly, given all the work on determining the incidence of IgH translocations, a consensus does not exist. The main problem is the varied results between studies in their detection of IgH translocations from a low of 23/55 (41.8%) in one study using metaphase FISH to a high of 43/45 (95.5%) in another study using interphase FISH[63,64]. In the largest cohort studied to date, Avet-Loiseau et al. detected IgH translocations in 477/653 (73%) of patients using interphase FISH[65]. Several factors influence the wide range of frequencies identified by different groups, including differences between FISH probes, selection of clonal plasma cells, cutoff levels, the enumeration of patients with chromosome 14 deletions, and the potential higher sensitivity of interphase FISH compared to metaphase

FISH[66]. At the very least 50% to 70% of myeloma patients have IgH translocations[67]. The incidence of translocations involving the IgL loci has not been extensively studied by FISH, but they are detectable by conventional cytogenetics in 8/370 (2.2%) of patients[46,66,68]. The non-immunoglobulin translocations involve a large variety of chromosomal loci but commonly involve either arm of chromosome 1, 6p10, 8q10, 8q24, and 11q13.

Although the presence of 14q+ marker chromosomes was initially reported in 1975 it took another nine years before Venti et al. identified the first partner chromosome[69-72]. The first IgH translocation identified was t(11;14)(q13;q32) as it is easily identified by conventional cytogenetics[72]. Even though 14q+ was identified as a recurrent abnormality in myeloma very few partner chromosomes were identified before 1996. The partner chromosomes are difficult to identify as most IgH translocations involve telomeric chromosomal segments. As a result, conventional cytogenetics often fails to identify the partner domain and equally problematic, often fails to identify a large number of karyotypically silent IgH translocations. In 1996, Bergsagel et al. published a ground breaking paper using a sensitive southern blot assay designed to detect IgH switch translocations in which they identified two of the most common IgH translocations; t(4;14)(p16;q32) and t(14;16)(q32;q23) for the first time[73]. Moreover, they identified several less common translocations; t(1;14)(p13;q32), t(8;14)(q24;q32), t(12;14)(q24;q32), and t(14;21)(q32;q22). Although several additional translocations; in particular t(6;14)(p25;q32), t(6;14)(p21;q32), and t(14;20)(q32;q11~q12) are suggested to be recurrent, no large cohort studies of these translocations are published[74-77]. Due to their recurrent nature, only t(4;14)(p16;q32), t(11;14)(q13;q32) and t(14;16)(q32;q23) have been studied in large cohorts[1,26,65,78,79]. Although these studies identified some limited frequency differences, they are largely similar to one another. The differences arise from several fundamental problems that only recently became apparent. The main problems are assay design and enumeration. Initially the IgH translocations were believed to be reciprocal events generating two derivative chromosomes. With this assumption most groups initially used assays designed to detect only one derivative[65,80]. Alternatively, some groups used well designed assays that could detect both derivative chromosomes but only enumerated cells with reciprocal rearrangements[81]. However, recently it became apparent that, at least for the recurrent translocations, approximately 30% of patients have unbalanced translocations[1,26,82]. Therefore, many of the initially reported frequencies are underestimates. In general the accepted frequencies in unselected patient populations for t(11;14), t(4;14), and t(14;16) are 15-20%, 12-15%, and 5%, respectively[67]. Given the occurrence of monosomy 14 in ~10% of patients and the high incidence of unbalanced translocations it is essential for assay design and enumeration techniques to be well thought out and justified.

Several groups have studied the clinical impact of the various recurrent IgH translocations. In the studies with large cohorts t(11;14) is predictive of a good prognosis while t(4;14) and t(14;16) predict for a poor prognosis[1,26,83]. Moreover, the recurrent IgH translocations are associated with a non-hyperdiploid karyotype[63,78]. Furthermore, t(4;14) and t(14;16) are associated with chromosome 13 deletions and deletions of the *TP53* locus[26,63,81,84]. Interestingly, t(4;14) and t(11;14) are associated with distinct plasma cell morphologies, immature/intermediate and CD20$^+$/lymphoplasmacytic, respectively [85-87].

Several basic problems require investigation now that several of the complexities associated with the recurrent translocations are known. First, at a maximum, only 40% of patients have one of the recurrent IgH translocations[26,65,79]. However, IgH translocations, detected by IgH break apart FISH assays, are detected in more than 70% of patients[65,79]. Therefore either a large number of translocations exist or potentially other recurrent translocations exist. These translocations are most likely cryptic and, like t(4;14), involve sub-telomeric regions, since they are not identified by conventional cytogenetics. Therefore, it is possible that unidentified recurrent translocations exist in patient populations, since most IgH translocations were initially identified in cell lines, but the malignant cells that form cell lines are most likely not representative of those in the general myeloma population. Furthermore, the incidence of the identified translocations is much higher in cell lines compared to patient populations highlighting their increased ability to proliferate in cell culture conditions. Thus several recurrent translocations may exist in patient samples that for a variety of reason do not support the development of a cell line and thus remain uncharacterized.

I.4 - The t(4;14)(p16;q32) Translocation

The second most common IgH switch translocation in multiple myeloma is t(4;14)(p16;q32)[26]. This translocation involves the sub-telomeric regions of chromosome arms 4p and 14q[88](Figure I.2). The translocation appears to be caused by illegitimate class switch recombination events[89]. This translocation highlights a unique feature of IgH switch translocations, which is the separation of the IgH mu and alpha enhancers onto the different derivative chromosomes (Figure I.2). Therefore, the expression of genes on either side of the 4p16 genomic breakpoints may be influenced by the respective IgH enhancer. This presents a fundamental problem in understanding the biology of t(4;14) myeloma as both *FGFR3* and *MMSET*, which are separated onto different derivative chromosomes, are potential target genes [88,90,91]. Furthermore, several other genes proximal to the common breakpoint region between *LETM1* and *MMSET* on chromosome 4 are proposed alternative target genes. These include; *TACC3*, *LETM1*, and *WHSC2*[1,92]. Moreover, at least in the case of t(14;16)(q32,q23) and t(14;20)(q32;q11~12) the IgH enhancers are capable of affecting the expression of genes located over 1 Mb away from the enhancer[77,93]. Therefore, it is essential to determine if t(4;14) dysregulates a single or many gene(s) and how those events contribute to myelomagenesis.

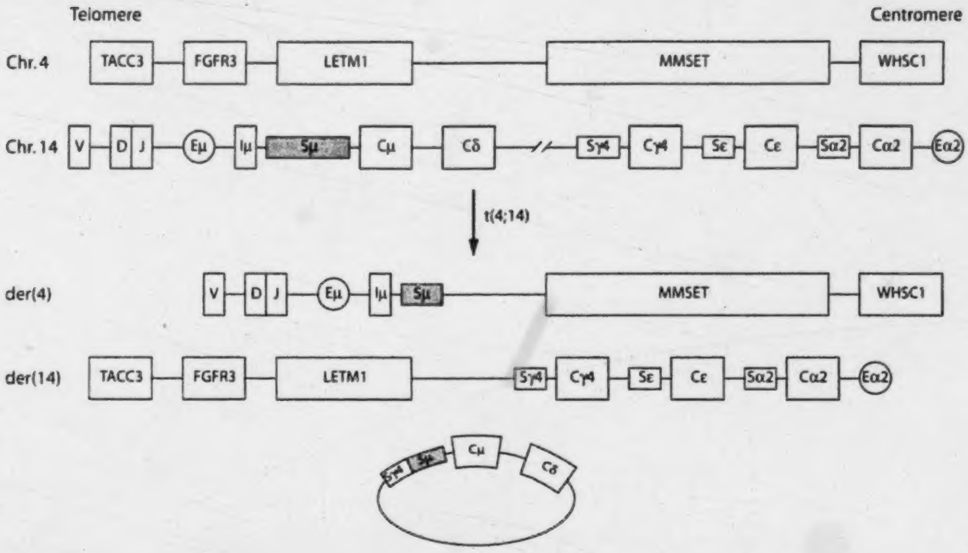

Telomere
Centromere
Chr. 4
TACC3
FGFR3
LETM1
MMSET
WHSC1
Chr. 14
V
D
J
Eμ
Iμ
Sμ
Cμ
Cδ
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
t(4;14)
der(4)
V
D
J
Eμ
Iμ
Sμ
MMSET
WHSC1
der(14)
TACC3
FGFR3
LETM1
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
Sγ4
Sμ
Cμ
Cδ

Figure I.2 – Basic Diagram of t(4;14)(p16;q23)

The 4p16 and 14q32 loci involved in t(4;14) are shown telomere to centromere. Furthermore, the resulting der(4) and der(14) chromosomes along with the excised "switch circle" from an illegitimate switch recombination resulting in t(4;14) are shown. As a result of the illegitimate switch recombination event the two immunoglobulin enhancers shown, Eμ and Eα2, are separated onto different chromosomes and potentially results in gene dysregulation on both derivative chromosomes. The excised switch circle should be lost shortly after the initial translocation event.

I.4.1 - The Initial Description of t(4;14)(p16;q32) in Multiple Myeloma

Analysis of multiple myeloma by conventional cytogenetics identified 14q32 translocations as common abnormalities[45,46,94]. However, using standard G-banding procedures the chromosomal partners involved could only be identified in half of the patients with this abnormality. Based on the assumption that the observed 14q32 translocation in human myeloma would occur in a similar mechanism as seen in mouse plasmacytomas, which involve IgH switch regions, Bergsagel et al. developed a comprehensive Southern blot assay that could detect legitimate and illegitimate IgH switch recombination events[73]. Using this strategy they showed that illegitimate switch recombination events are common in cell lines and primary tumour samples. Furthermore, 7 of 8 cell lines without karyotypically identified 14q32 translocations harbored illegitimate switch recombinations, suggesting IgH switch translocations were even more prevalent than initially proposed. Within the panel of samples analyzed, 4p16.3 was identified as a recurrent partner domain. However, at the time, and unlike t(11;14), the most common recurrent IgH switch translocation identified by cytogenetics, no dysregulated target gene was identified.

The first comprehensive analysis of t(4;14) was published in July 1997 by Chesi et al[88]. This study represented a detailed analysis of t(4;14) with a more specific Southern blot assay, which included probes specific for 4p16. The continued use of the Southern blot assay was necessary because the translocation is karyotypically silent, since the involved regions, 4p16.3 and 14q32, exist in sub-telomeric domains of their respective chromosomes. Sequencing of the cloned switch translocation breakpoints from 4 t(4;14) positive cell line and one patient sample confirmed the presence of 4p16.3 associated sequences. In all cases the identified segment of 4p16.3 was centromeric of *FGFR3*. Therefore, *FGFR3* was proposed to be the target gene of t(4;14) as the presence of the 3' alpha IgH enhancer(s) and FGFR3 on der(14) was predicted to result in FGFR3 overexpression. To test this hypothesis a panel of cell lines and patient samples were screened for *FGFR3* expression by RT-PCR. This showed that 4 of 21 cell lines, three of which were previously described as t(4;14) positive by southern blot, overexpressed *FGFR3*. Interestingly, one t(4;14) cell line identified by southern blot, JIM3, did not express *FGFR3* at a detectable level. However, the original tumour sample from which the cell line was derived was shown to express *FGFR3*. The additional *FGFR3* expressing cell line, UTMC-2, was subsequently shown to be t(4;14) positive by Southern blot. *FGFR3* expression was detectable in 4 of 10 patient samples of which 3 were confirmed as t(4;14) positive by Southern blot. To strengthen the argument that *FGFR3* was the target gene of t(4;14) the *FGFR3* mutation status of *FGFR3* expressing samples was determined. Two cell lines and one patient sample, all t(4;14) positive by Southern blot, expressed a mutated allele of *FGFR3*, known to cause hyper-activation of the FGFR3 signaling cascade. All three mis-sense mutations were previously identified in the autosomal dominant skeletal disorders (hypochondroplasia, achondroplasia, thanatophoric dysplasia type I and II) and therefore were viewed as somatic mutation events since none of the samples originated from a person with a skeletal disorder [95].

Subsequent to the characterization of t(4;14) in cell lines by Chesi et al. the translocation was independently characterized in a Southern blot screen of primary myeloma patient samples by Richelda et al[96]. However, their strategy used different restriction enzymes and probes. Using this strategy, three t(4;14) positive samples were identified and all three expressed wild-type *FGFR3* transcripts.

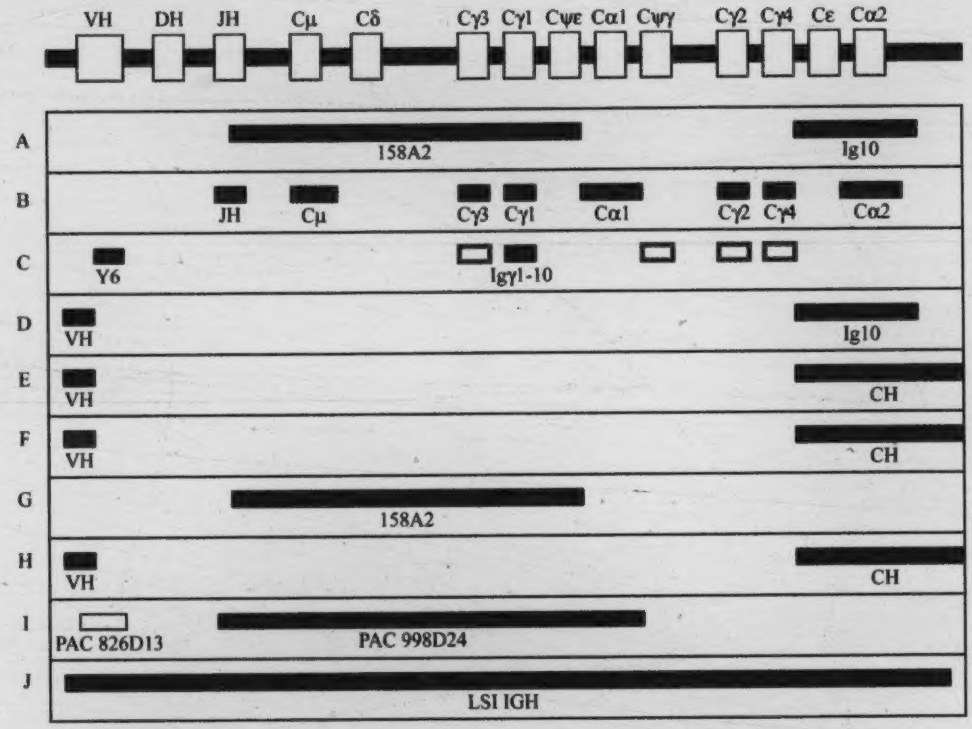
I.4.2 - Improving the Characterization of t(4;14)(p16;q32)

The initial characterization of t(4;14) required an elegant and intuitive assay design, but the Southern blot assays have several limitations. First, the genomic regions involved in both legitimate and illegitimate switch recombinations are large and thus require numerous probes to obtain complete coverage. Second the restriction endonuclease used to digest genomic DNA may generate bands that do not transfer efficiently or result in recombined bands that co-migrate with the genomic bands. Thus, the Southern blot assays underestimated the occurrence of t(4;14) due to the low sensitivity of the assay.

The first improvement was the use of metaphase FISH. The first published assay was a two colour assay using a chromosome 14 painting probe and a cosmid probe specific for *FGFR3*[88]. Using this strategy t(4;14) positive samples are identified by the co-localization of the chromosome 14 painting probe and the *FGFR3* probe on der(14). The use of SKY appeared promising, however, this strategy fails to identify t(4;14) positive samples as both regions are sub-telomeric[68,97,98]. Ultimately, strategies requiring metaphase preparations are not commonly used as they are difficult and inconsistently generated from primary myeloma samples.

The second improvement in the detection of t(4;14) was the use of RT-PCR to detect hybrid transcripts and the dysregulation of *FGFR3* transcription caused by the translocation. These new methods were pioneered by Chesi et al.[88,91], and they detect RT-PCR products only in patients with the translocation. The first published assay was the detection of *FGFR3* transcripts[88], made possible by the observation that *FGFR3* transcripts are below the level of detection in normal plasma cells, BMMC, and peripheral blood mononuclear cells (PBMC). As a direct result of t(4;14), the *FGFR3* locus is brought into close proximity of the 3' IgH enhancers resulting in a high level of transcription from the *FGFR3* locus that is easily detected by RT-PCR. Therefore, cell lines and patient samples can be screened for *FGFR3* expression and positive results identifying t(4;14) positive samples. Several groups have used this method to identify t(4;14) positive patients as the specificity is quite high with few t(4;14) negative samples having detectable *FGFR3* transcripts. However, the sensitivity of the assay is compromised by the fact that not all t(4;14) positive samples express *FGFR3*[1,82,88,91,99,100]. The second published RT-PCR assay detects novel hybrid transcripts created by the translocation[91]. As a direct result of t(4;14) two different sets of hybrid transcripts can be detected which are transcribed from either the der(4) or der(14) chromosomes. On the der(4) chromosome transcription events originating within the JH and Iμ loci result in spliced transcripts containing the respective IgH element and the *MMSET* gene, termed IgH-MMSET hybrid transcripts. Alternatively on the der(14) transcription events originating within the 5' untranslated region (UTR) of the *MMSET* gene result in spliced transcripts containing the *MMSET* gene and an IgH constant gene, termed MMSET-IgH transcripts. Moreover, the relative breakpoint location on chromosome 4 can be predicted based on the product sized detected in the der(4) IgH-MMSET hybrid transcript assays Both of these assays are highly specific but they have different levels of sensitivity[1,91,101]. The IgH-MMSET hybrid transcripts are detected in the large majority of t(4;14) positive samples with only one t(4;14) positive cell line (LP-1) and one t(4;14) positive patient identified by FISH being negative to date[82,91]. However, less than 50% of t(4;14) positive samples have detectable MMSET-IgH hybrid transcripts[1,91]. When used in combination these RT-PCR assays are capable of identifying nearly all t(4;14) positive samples identified by other means[82,91,101].

The third improvement in the detection of t(4;14) was the use of interphase FISH. There are a variety of interphase FISH strategies used by different groups. For the most part they are evolving with our understanding of t(4;14) biology (Figure I.3). The first publication of this method was by Avet-Loiseau et al.[102]. This initial design used a mix of two probes that cover the entire IgH constant region in combination with a probe that flanks the *FGFR3* gene and exon 1 of *MMSET*. Similarly Finelli et al. used a mixed panel of probes that cover all of the IgH constant regions except IgD and IgE in combination with a probe that flanks *FGFR3* and *WHSC2*[103]. In both cases the assays were validated on two known t(4;14) positive cell lines, NCI-H929/OPM-2 and KMS-11/OPM-2 respectively. However, both protocols are flawed because the co-localized probes that indicated the presence of t(4;14) are located on der(14). Finelli et al. validated the positive samples with a JH probe that could detect the der(4) chromosome, but this was only used as a validation of the der(14) result and thus was not performed on the der(14) negative samples. Subsequent to these initial interphase FISH papers the strategies have evolved and in general improved. The new strategies used by the research groups of John Shaughnessy and Rafael Fonseca along with the commercially available LSI IGH/FGFR3 probe set from Vysis appear to be the most comprehensive strategies, which given current knowledge would identify all t(4;14) positive patients. Currently interphase FISH with the proper probe design on plasma cells selected by CD138 expression, light chain restriction, or morphology is the accepted gold standard method of detecting t(4;14).


VH
DH
JH
Cμ
Cδ
Cγ3
Cγ1
Cψε
Cα1
Cψγ
Cγ2
Cγ4
Cε
Cα2
A
158A2
Ig10
B
JH
Cμ
Cγ3
Cγ1
Cα1
Cγ2
Cγ4
Cα2
C
Y6
Igγ1-10
D
VH
Ig10
E
VH
CH
F
VH
CH
G
158A2
H
VH
CH
I
PAC 826D13
PAC 998D24
J
LSI IGH

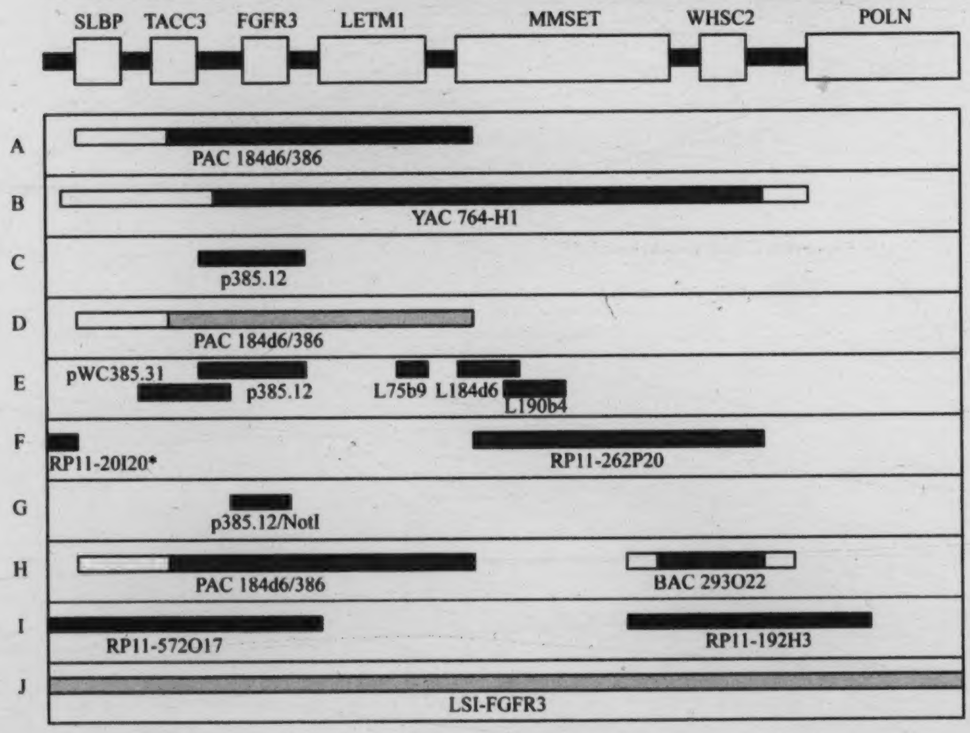

SLBP
TACC3
FGFR3
LETM1
MMSET
WHSC2
POLN
A
PAC 184d6/386
B
YAC 764-H1
C
p385.12
D
PAC 184d6/386
E
pWC385.31
p385.12
L75b9
L184d6
L190b4
F
RP11-20I20*
RP11-262P20
G
p385.12/NotI
H
PAC 184d6/386
BAC 293O22
I
RP11-572O17
RP11-192H3
J
LSI-FGFR3

Figure I.3 – FISH Strategies Used by Different Groups to Detect t(4;14)

A general map of the two genomic regions involved in t(4;14) is shown in a telomeric to centromeric orientation. Characterized genes or IgH regions are noted with open boxes and the corresponding name is noted above each box. The relative location of each FISH probe based on accurate mapping information is noted by solid boxes. When the flanking sequence of a probe is not specifically mapped, the potential region contained in the probe is represented as an open box. When possible the colour of the probes used in each referenced publication is indicated by the annotated colour. (A) The first FISH strategy employed by the group of Herve Ayet-Loiseau on behalf of the Intergroup Francophone du Myelome (IFM) to detect t(4;14)[79,84,102,104]. The centromeric portion of the IgH locus was probed with the cosmid Ig10 (Dr. Rabbits, Cambridge, UK) and BAC 158A2 (Dr. Batzer, New Orleans, LA). The cosmid Ig10 is predicted to flank Cα2 and Cε though the specific sequence of the flanking regions is not available. The BAC 158A2 is predicted to flank the JH region and Cψε, but again no specific flanking sequence is available. While 4p16.3 is probed with the PAC 184d6/386 (Genome Systems), which was described by Chesi et al.[88]. The ends of PAC 184d6/386 were sequenced[88], however, the sequence is not deposited in NCBI or other sequence repositories and therefore the specific locations are unknown. But based on available information the centromeric end of the probe sits between exons 2c and 2d of *MMSET* while the telomeric end is not specifically mapped, but does include the entire *FGFR3* locus. (B) The FISH strategy used by the group of Antonio Neri (Milan, Italy) to detect t(4;14)[101,103]. The IgH region was probed with a panel of plasmids containing sequences specific for each constant region and the JH region. The gamma constant regions (1-4) are fragments of 7, 4, 7, 6 kb generated from a BamH1/HindIII digest. The alpha constant regions (1 and 2) are fragments of approximately 18 kb generated from a BamH1 digest. The mu constant region fragment generated by a HindIII digest is 10 kb in length. The JH fragment generated by a BamH1/HindIII digest is a 6.6 kb in length. This latter probe is labeled in grey as it was not used to screen patients but to detect the der(4) on patients with an identified der(14)[103]. The 4p16 region was probed with YAC 764-H1 from the Centre d'Etude du Polymorphisme Humain (CEPH)(Paris, France). No specific sequence information is available but the YAC is approximately 400 kb in length and based on available mapping, contains sequence centromeric of *WHSC2* and telomeric of *FGFR3*. (C) Strategy employed by Nakazawa et al. to detect t(4;14)[99]. The telomeric VH region was probed with YAC Y6 (Dr. Matsuda, Kyoto, Japan), which is approximately 310 kb in length and flanks the VH segments VH3-64 and VH3-32P at its telomeric and centromeric ends, respectively[105]. This sequence lacks approximately 170 kb of the telomeric IgH sequence that includes the 14q telomere[106]. The centromeric region was probed with a bacteriophage clone Igγ1-10 (Dr. Matsuda, Kyoto, Japan) that is approximately 10 kb in length and cross hybridizes with the other gamma segments (Noted by grey boxes). The cosmid pC385.12 (M.R. Altherr, Los Alamos, NM) was used to probe 4p16.3. Specific sequence information is not available but the cosmid contains the entire *FGFR3* locus and is consistently represented as being approximately 40 kb in length[88,99]. (D) The second FISH strategy used by the group of Herve Avet-Loiseau on behalf of the IFM[65,83,85]. This updated strategy utilized the cosmid yIgH6-9 to identify the VH region and the cosmid Ig10 to identify the centromeric portion of the IgH constant region. The cosmid yIgH6-9 is approximately 35 Kb in length and flanks the

VH segments VH7-77 and VH3-74 at its telomeric and centromeric ends, respectively, which are telomeric of VH3-64[107]. This cosmid was generated from the YAC yIgH6 that includes the 14q telomere and VH3-64. To identify 4p16 the PAC 184d6/386 was used. (E) Set of FISH probes used by Fonseca et al. to detect t(4;14) in MGUS patients[108]. The VH and CH region are identified with the cosmid yIgH6-9 and CH BAC (Genome Systems), respectively. No specific sequence information is available for the CH BAC, but the initial description of this probe by Gabrea et al. indicates that it flanks Cε and the Cα2 enhancer[109]. A panel of cosmid probes, L190b4, L184d6, L75b9, pC385.12, and pWC385.31 that flank known t(4;14) breakpoints were used to identify 4p16. Cosmid L190b4 (NCBI Accession Z68276) is 29 kb in length and flanks exons 4a and 8 of *MMSET*. Cosmid L184d6 (NCBI Accession Z49236) is 28 kb in length and flanks exons 2a and 4 of *MMSET*. Cosmid L75b9 (NCBI Accession Z69653) is 5 kb in length and sits within intron 1 of *LETM1*. No specific sequence information is provided for pWC385.32, however, based on the referenced publication it would represent sequence telomeric of *FGFR3*. (F) FISH probes used by Santra et al. to identify the dissociation of VH/CH and *FGFR3*/*MMSET* in t(4;14) positive myeloma cases lacking *FGFR3* expression[82]. The VH and CH probes are the cosmid yIgH6-9 and CH BAC, respectively. The BAC RP11-20I20 (NCBI Accession AC092535) is 190 kb in length and covers the *SPON2* and *MAEA* genes at its telomeric and centromeric ends, respectively. Importantly this probe, does not include *FGFR3* sequences as suggested by Santra et al. as it sits over 438 kb telomeric of the *FGFR3* locus, and thus is noted with an asterisks as the sequence is not present on the map. The BAC RP11-262P20 (NCBI Accession AL132868) is 193 kb in length and covers 4p16 sequence telomeric of

MMSET exon 2a and centromeric of *WHSC2*. (G) FISH probes used at the Princess Margaret Hospital/University Health Network (Toronto, Ont) to detect t(4;14) by cytoplasmic immunoglobulin enhanced FISH (cIg-FISH)[27,80,110,111]. The 158A2 BAC is used to probe the telomeric half of the IgH constant region. A 22 kb *FGFR3* segment is used to probe the *FGFR3* region. The *FGFR3* fragment is generated by a NotI digestion of p385.12 (Hong Chang, Pers. Comm.). (H) The FISH probes used by the group of Rafael Fonseca at the Mayo Clinic in recent years[26,78,81]. The telomeric and centromeric regions of the IgH locus are probed with the cosmid yIgH6-9 and CH BAC (Genome Systems), respectively. The telomeric side of the 4p16.3 breakpoints is probed with PAC 184d6/386 and the centromeric side of the known breakpoints is probed with BAC 293O22 (Rafael Fonseca, Pers. Comm.) from Incyte Genomics, that flanks cosmid L96a2 (NCBI Accession Z68165). Cosmid L96a2 flanks exon 24 of *MMSET* and the *WHSC2* gene. (I) The improved FISH strategy used recently by the group of Antonio Neri (Milan, Italy) to detect both derivative chromosomes[112]. FISH probes for the IgH constant and variable regions are PAC clones 998D24 and 826D13 respectively. The PAC 998D24 flanks the JH region and Cα1 (PAC end sequences, NCBI Accession AZ579060 and AZ579059). This map localization is discrepant with that proposed by Fabris et al. since their initial characterization of this probe suggested that it contained Cα sequence but not JH sequence as detected by a PCR screen[113]. No specific information is available or provided to characterize the map location of PAC 826D13. FISH probes specific for the telomeric and centromeric sides of known 4p16 breakpoints are BACs RP11-572O17 (NCBI Accession AC016773) and RP11-192H3 (BAC end sequences, NCBI Accession AQ412237 and AQ412239) respectively. RP11-572017 is

189 kb in length and flanks *SLBP* and *FGFR3* at its telomeric and centromeric ends respectively. RP11-192H3 is 173 kb in length and includes exon 18 of *MMSET* and the telomeric half of *POLN.* (J) The LSI® IGH/FGFR3 dual color, dual fusion translocation probe set available commercially from Vysis (Downers Grove, Ill)[114].

The fourth improvement in the detection of t(4;14) was the use of global gene expression profiling with microarray platforms. When gene expression profiles are analyzed for the expression of *FGFR3* and *MMSET* transcripts the t(4;14) positive patients are easily identified[82,112,115,116]. In the majority of patients *FGFR3* and *MMSET* are overexpressed uniquely in the t(4;14) positive samples, however, approximately 30% of the t(4;14) positive patients do not express *FGFR3*[82,112]. This can present a limitation as the expression of *MMSET* in t(4;14) negative samples using gene expression profiling (GEP) is controversial, as discussed below. In one publication Dring et al. found 5 of 24 t(4;14) negative samples expressed *MMSET* at levels comparable to those observed in t(4;14) positive samples[117]. In another recent study Fabris et al. found one t(4;14) negative patient expressing *MMSET* at levels comparable to those seen in t(4;14) positive patients[112]. This patient presented a unique cytogenetic finding as the genetic locus containing *FGFR3* and *MMSET* was transferred from chromosome 4 to an unidentified chromosome. Additionally, 13 of the 32 remaining t(4;14) negative cases had *MMSET* expression levels above the cutoff level. However, in another study Santra et al. found an approximate 16 fold increase (Mean, Range; 691, 298-1179 versus 11230, 4098-24 885) in the expression of *MMSET* in t(4;14) positive/FGFR3 negative patients compared to the t(4;14) negative patients[82]. With this dichotomy in mind, a larger set of genes which are differentially expressed in positive and negative patients is needed to prospectively characterize the t(4;14) status of a patient. Using a supervised analysis Dring et al. identified 127 genes differentially expressed between t(4;14) positive and negative samples[117]. However, this cohort did not include any t(4;14) positive/FGFR3 negative patients and the differentially expressed genes were not tested on an independent or unanalyzed data set to determine if they could accurately identify the t(4;14) status of a sample. Therefore, a well validated analysis is needed before the t(4;14) status of a sample can be determined with a high degree of accuracy using this approach.

I.4.3 - The Incidence of t(4;14)(p16;q32) in Multiple Myeloma

The initial discovery of t(4;14) was made in a panel of cell lines of which approximately 25% are t(4;14) positive[88]. Subsequently a large number of studies have looked at the incidence of t(4;14) in patient cohorts using Southern blots, metaphase FISH, RT-PCR, interphase FISH, and GEP (Table I.2). Based on all of this work the currently accepted frequency of t(4;14) in overt multiple myeloma is approximately 15%. Some studies observed frequencies in the range of 10-12%, however, most of these strategies used interphase FISH assays specific for the der(14), which we now know is commonly lost[1,26,82,112], or the studies were biased in their counting towards balanced translocations. With most new studies the FISH strategies are designed to detect both derivative chromosomes and the scoring bias towards balanced translocations has been removed. Alternatively, the IgH-MMSET assays specific for the der(4) chromosome are commonly used. These two strategies appear to identify the large majority of patients with only 2 patients and 1 cell line being scored differently between the two assays[82,91,118]. When the der(4) IgH-MMSET hybrid transcript assays are used to screen patients, the incidence of t(4;14) is 13.9%. Excluding the initial work of Malgeri et al. no one has compared interphase FISH and the der(4) IgH-MMSET hybrid transcript assays[101]. In that initial paper, a perfect concordance between the two assays was observed, but they used a der(14) specific interphase FISH assay and a limited number of patients (n=53). It is currently necessary that a large group of patients be screened with one of the improved FISH assays (that detect both derivatives and accurately score unbalanced translocations) and the IgH-MMSET assays to determine a highly accurate incidence rate and evaluate the relative sensitivity of each assay.

Table I.2 – The Incidence of t(4;14) in Multiple Myeloma Patients

	Assay	Multiple Myeloma			MGUS	
		Patients	t(4;14) (%)	t(11;14) (%)	Patients	t(4;14) (%)
Chesi (1997)[88A]	RT-PCR (FGFR3) der(14)	10	3			
Avet-Loiseau (1998)[102]	FISH der(14)	135	17 (12.6)	23 (17.0)		
Finelli (1999)[103]	FISH der(14)	30	5 (16.7)			
Avet-Loiseau (1999)[84G]	FISH der(14) CD138$^+$				100	2
Malgeri (2000)[101C]	RT-PCR, FISH der(14)	53	11 (20.8)		16	1
Nakazawa (2000)[99E]	FISH der(14) & RT-PCR (FGFR3)	45	8 (17.8)			
Avet-Loiseau (2001)[104I]	M-FISH & FISH der(14)	40	5 (12.5)	13 (32.5)		
Fonseca (2001)[81K]	cIg-FISH der(4&14)	155	16 (10.3)			
Sibley (2002)[119]	RT-PCR	67	7 (10.4)		13	2

Fonseca (2002)[108]	cIg-FISH 47% $CD138^+$ der(4&14)				56	5
Avet-Loiseau (2002)[65H]	FISH der(14) $CD138^+$	715	74 (10.3)	120 (16.8)	168	4
Moreau (2002)[83]	FISH der(14) $CD138^+$	168	22 (13.1)	26 (15.5)		
Rasmussen (2002)[120]	qRT-PCR (FGFR3) der(14)	110	16 (14.5)			
Soverini (2002)[118]	qRT-PCR (FGFR3) der(14)	78	10 (12.8)			
Keats (2003)[1]	RT-PCR	208	31 (14.9)		52	1
Santra (2003)[82]	GEP, RT-PCR, cIg-FISH der(4&14)	178	32 (18.0)			
Fonseca (2003)[26]	cIg-FISH der(4&14)	332	42 (12.7)	53[F] (15.8)		
Rasmussen (2003)[121L]	RT-PCR, qRT-PCR (FGFR3) der(14)	40	3 (7.5)		20	0
Chang (2004)[80]	cIg-FISH der(14)	120	15 (12.5)	16[D] (12.8)		
Chang (2005)[110J]	cIg-FISH der(14)	14	2 (14.3)	5 (35.7)		
Rasmussen (2005)[32]	RT-PCR	54	7 (13.0)			
Fabris (2005)[112M]	RT-PCR, GEP, FISH der(4&14)	45	6 (13.3)	10 (22.2)		
Keats (2005)[122N]	RT-PCR	304	43 (14.1)		112	2
Totals (All)		2670	340 (12.7)	256/1533 (16.7)	485	16 (3.3)
Totals Only der(14) Tested		1465	177 (12.1)		268	6 (2.2)

Totals der(4) Tested		1228	167 (13.6)		217	10 (4.6)

The screening results from papers using a Southern blot assay are not included as the frequencies are substantially below those reported in the subsequent studies with more accurate assays. Unless otherwise noted when RT-PCR is noted as the screening assay this reflects the use of the der(4) IgH-MMSET hybrid transcript assays.

A) Samples were screened by *FGFR3* expression, which identified 4 positive patients, however, one patient was not reported as being positive for t(4;14) in this paper and is excluded.

C) The patient cohort includes 23 previously studied patients of which 4 are t(4;14) positive. Therefore only 30 additional patients are included in total number of patients screened and only 7 additional t(4;14) positive samples are counted in the total number of t(4;14) positives.

D) In this study 128 patients were studied, however, only 120 of them were assayed for t(4;14) and only 125 were assayed for t(11;14).

E) In this study 7 of 45 were t(4;14) positive by FISH. Interestingly, one patient was FISH negative but did overexpress *FGFR3*. Since the FISH strategy would miss a translocation involving the Cα2 switch region, this patient has been included as a t(4;14) positive.

F) In this study 351 patients from the Eastern Cooperative Oncology Group (ECOG) E9487 trial were studied, however, only 332 were assayed for t(4;14) and only 336 were assayed for t(11;14).

G) This cohort includes both MGUS and sMM patients diagnosed based on the criteria set forth by Kyle et al. [123,124].

H) The MGUS cohort is composed of MGUS and sMM patients while the MM cohort contains both frank MM and PCL patients.

I& J) All patients included in the studies were diagnosed with PCL not MM.

K) The 55 MGUS samples are not noted as I believe they correspond to the same patients listed in Fonseca (2002)[108].

L) One patient is *FGFR3* positive but negative for IgH-MMSET hybrid transcripts. This patient is not included as the mechanism of *FGFR3* expression was not determined by alternative means.

M) Includes all patients studied; 39 MM and 6 PCL.

N) The original cohort of 208 MM and 52 MGUS are included within this cohort and therefore only 96 additional MM (12 new positives) and 60 MGUS (1 new positive) are included in the totals.

I.4.4 - The Significance of t(4;14)(p16;q32) in Multiple Myeloma

The identification of recurrent IgH translocations in multiple myeloma prompted the hypothesis that these genetic events may predict prognosis or response to different therapies. Several initial studies attempted to determine if a correlation between prognosis and t(4;14) existed, however, they were small studies with limited power. The first study to comment on the clinical outcome associated with t(4;14) actually showed no significant difference between t(4;14) positive and negative patients[120]. However, this study identified t(4;14) based solely on the expression of *FGFR3* and only looked at 76 patients. Subsequent to that initial report three independent groups (Herve Avet-Loiseau, France; Linda Pilarski, Canada; and Rafael Fonseca, United States) presented data at the

43rd annual meeting of the American Society of Hematology (Dec 2001) showing t(4;14) predicted for a poor clinical outcome. The first publication by Moreau et al. showed that t(4;14) as detected by interphase FISH (probes specific for der(14)) predicted for a poor overall outcome in patients treated with high dose therapy (HDT) (principally, single autologous transplants) with a median survival of 32.8 months[83]. The second publication, by our group, showed that t(4;14) detected by RT-PCR for der(4) IgH-MMSET hybrid transcripts predicted for a poor overall outcome, median overall survival of 21.1 months[1]. Unlike the study of Moreau et al. we found a differential response to therapy between the t(4;14) positive and negative patients. However, our cohort was not treated in a uniform fashion and since this was a retrospective analysis, data was not available for all patients. Similar to the study of Rasmussen et al. we found that *FGFR3* expression, which only identifies ~70% of t(4;14) positive patients, did not predict for a differential outcome. It is interesting that Moreau et al. saw a significant difference in outcome since their interphase FISH assay detects the der(14), which in theory, should detect the same population of patients identified by *FGFR3* expression. However, as suggested by Santra et al.[82], some der(14) positive patients may lose *FGFR3* expression by a yet undetermined mechanism. The third publication by Fonseca et al. showed that t(4;14) detected by cIg-FISH (using probes for both derivatives and scoring unbalanced translocations) predicted for a poor overall outcome in patients treated with conventional chemotherapy (ECOG trail E9486, $\text{VBMCP}^{+/-}$INF-α2 and early cyclophosphamide intensification[125]), median overall survival 26 months[26]. Similar to the study of Moreau et al. this study did not observe any differences in the response to therapy in t(4;14) positive or negative patients. These initial results were confirmed recently by Chang et al. who studied 120 consecutive myeloma patients treated with 4-5 cycles of VAD followed by high dose melphalan and autologous peripheral blood stem cell (PBSC) transplant[80]. The t(4;14) positive patients were detected by cIg-FISH (using probes specific for der(14), though not all potential der(14) would be detected due to a poorly selected IgH constant region probe) and as previously observed the translocation predicted for a poor overall outcome with a median survival of ~27.9 months from diagnosis[iv]. Currently, t(4;14) is accepted as one of the worse prognosis groups in myeloma and prospective clinical trials for this subgroup are ongoing (Rafael Fonseca, Pers. Comm).

Recurrent genetic markers can be useful prognostic indicators but are equally, or even more importantly, good predictors of treatment efficacy. Two good examples of this phenomenon occur in acute myelogenous leukemia (AML) and chronic myeloid leukemia (CML). In AML, which has numerous morphological and genetic subtypes, the t(15;17)(q22;q21) translocation (PML-RARα) is the principle genetic characteristic of acute promyelocytic leukemia (French, American, British (FAB) subtype, AML M3 or M3v). This subgroup is effectively treated with all-trans retinoic acid (ATRA), while in the other subgroups, ATRA is ineffective. Similarly, in CML patients with t(9;22)(q34;q11), which creates the BCR-ABL hybrid, treatment with STI571/imatinib mesylate/Gleevec is highly effective. Unfortunately, to date, no therapy has been identified that works better, or even equally well, in t(4;14) patients.

[iv] In the actual manuscript, survival was presented as time post transplant not from diagnosis. The median survival post transplant was 18.3 months and the median time to transplant after diagnosis was 9.6 months.

I.5 - The Genes Associated with t(4;14)(p16;q32)

The chromosome bands 4p16 and 14q32 are the genetic loci involved in t(4;14). The breakpoints on chromosome 14 occur principally in the IgH switch regions[89]. The surrounding region is rich in coding elements required to produce an immunoglobulin heavy chain but other genes are extremely rare in this locus. The breakpoints on chromosome 4 occur within an evolutionary conserved paralogous chromosomal segment[126-128]. This chromosomal segment is highlighted by several genes with paralogs at 4p16, 5q35, 8p11 and 10q26. One of the gene families present at each chromosomal segment is the fibroblast growth factor receptor gene family. Numerous genes proximal to fibroblast growth factor receptor (FGFR) family members are conserved in two or three of the paralogous segments including members of the TACC, LETM, NSD, BAG, RGS, GRK, MXD, and PLEKHA families. Of these conserved genes, *FGFR3* was the initially proposed target gene of t(4;14) and *TACC3*, *LETM1*, and *WHSC1/MMSET/NSD2* are proposed alternative target genes[1,35,88,90-92,100,129,130]. To date, all of the cloned t(4;14) breakpoints are in a 65 kb region centromeric of *FGFR3* that includes portions of *LETM1* and *WHSC1/MMSET/NSD2*. Therefore, in most patients *TACC3*, *FGFR3*, and *LETM1* are transferred to the der(14) chromosome while *MMSET* stays on the der(4) chromosome. The *WHSC2* gene, which is immediately centromeric of *MMSET* is another proposed target gene of t(4;14)[1]. Interestingly, this gene does not have a paralog on one of the other paralogous chromosomal segments. The FGFR, TACC, and NSD paralogs have evolved in similar fashions (Figure I.4), suggesting that an initial duplication of the ancestral segment generated the segment present on chromosome 5q and subsequent duplications generated the segments present on chromosomes 4p, 8p, and 10q.

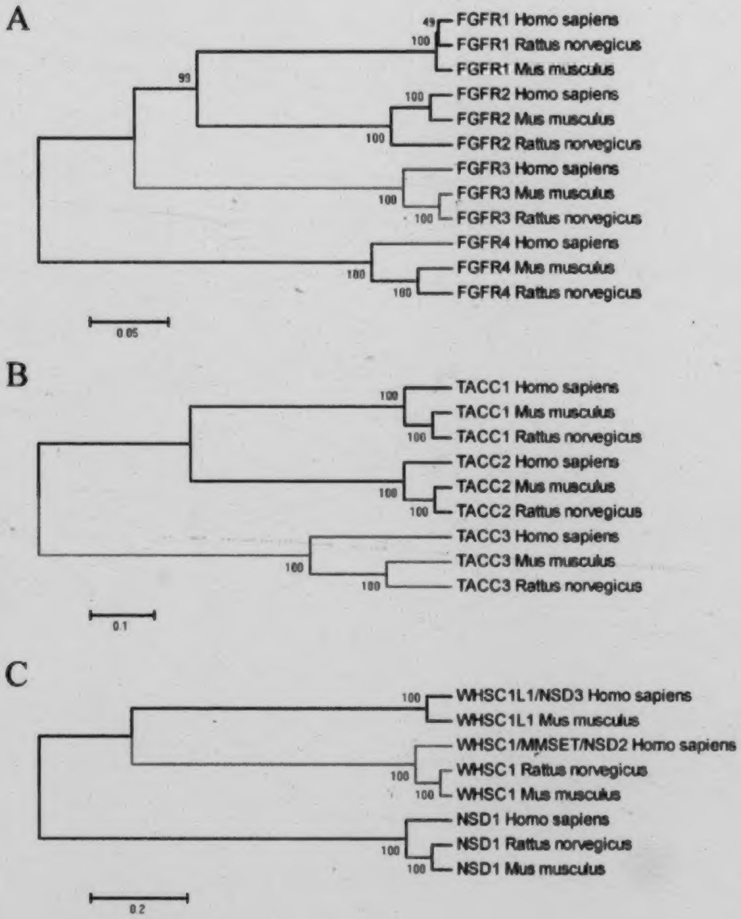

A
49
FGFR1 Homo sapiens
100
FGFR1 Rattus norvegicus
FGFR1 Mus musculus
99
100
FGFR2 Homo sapiens
FGFR2 Mus musculus
100
FGFR2 Rattus norvegicus
FGFR3 Homo sapiens
100
FGFR3 Mus musculus
100
FGFR3 Rattus norvegicus
FGFR4 Homo sapiens
100
FGFR4 Mus musculus
100
FGFR4 Rattus norvegicus
0.05
B
100
TACC1 Homo sapiens
TACC1 Mus musculus
100
TACC1 Rattus norvegicus
TACC2 Homo sapiens
100
TACC2 Mus musculus
100
TACC2 Rattus norvegicus
TACC3 Homo sapiens
100
TACC3 Mus musculus
100
TACC3 Rattus norvegicus
0.1
C
100
WHSC1L1/NSD3 Homo sapiens
WHSC1L1 Mus musculus
WHSC1/MMSET/NSD2 Homo sapiens
100
WHSC1 Rattus norvegicus
100
WHSC1 Mus musculus
NSD1 Homo sapiens
100
NSD1 Rattus norvegicus
100
NSD1 Mus musculus
0.2

Figure I.4 – Phylogeny of the FGFR, TACC, and NSD paralogs

UPMGA phylogeny trees were constructed with the MEGA 2.1 software package (www.megasoftware.net) from ClustalW multiple sequence alignments (www.ebi.ac.uk/clustalw) of human, mouse, and rat homologues [131,132]. Trees were generated using the amino acid poisson correction model and complete deletion of gaps/missing data. The accuracy of each tree was tested by 500 replication Bootstrap tests. The branch containing each human chromosome is noted with a unique colour, chromosome 4 (red), chromosome 5 (green), chromosome 8 (black), and chromosome 10 (blue). (A) FGFR1-4 family members. (B) TACC1-3 family members. (C) NSD1-3 family members.

I.5.1 - TACC3/ERIC-1

TACC3 is a strong candidate t(4;14) target gene. We and others proposed the involvement of *TACC3* based on several factors[1,92,100]. First, *TACC3* is only 50 kb telomeric of *FGFR3* and thus would be present on the der(14) created by t(4;14)(Figure I.4 & I.2). Second, *TACC3* is in the same transcriptional orientation as *FGFR3*, making the likelihood of *TACC3* dysregulation by the strong 3' IgH alpha enhancers, as seen for *FGFR3*, very likely. Third, the *TACC3* paralog *TACC1* is present in the 8p11 breast cancer amplicon and overexpression of TACC1 in an NIH 3T3 transformation assay is transforming[133]. Furthermore, the remaining paralog, *TACC2*, is implicated in breast cancer and may act as a tumour suppressor[134]. Fourth, the transforming effect of *TACC1* may relate to its function in stabilizing the centrosome. The ability of TACCs to stabilize centrosomes is tightly linked to their expression level as overexpression or underexpression can cause defects. So the potential exists for *TACC3* to be transforming when overexpressed by the strong 3' alpha enhancers or in a haplo-insufficient situation when the der(14) is lost.

To further support the potential involvement of *TACC3* in t(4;14) Still et al. found that by northern blot analysis, all cell lines tested except Raji expressed high levels of

TACC3[92]. Moreover, it appears as if *TACC3* expression is related to cell division as all of the *TACC3* positive tissues, identified by Northern blot, are proliferative[135]. Moreover, the *in vitro* stimulation of murine T and B cells confirmed the cell cycle regulated expression of *Tacc3*[136]. Initially, *Tacc3* was identified as the only TACC expressed in murine hematopoietic cells by northern blot[136], however, a similar analysis of human tissue by qRT-PCR identified *TACC1* and *TACC2* expression in the thymus, and *TACC1* expression in peripheral blood leukocytes[135].

The exact localization of TACC3 in interphase cells is questionable. Some groups find it in the nucleus and cytoplasm while others find it in the cytoplasm and peri-nuclear aggregates[136-138]. Importantly, unlike TACC2, neither TACC1 nor TACC3 are associated with centrosomes during interphase[137]. However, during mitosis TACC3 is strongly localized to a diffuse region around the centrosomes and to the mitotic spindle[136,137,139]. Unlike *Tacc2*, which has no phenotype[140], a homozygous deletion of murine *Tacc3* is embryonic lethal in mid gestation[136]. The embryonic lethal phenotype can be reversed in a p53 deficient background. Interestingly, the rate of tumour development in these mice is not increased compared to p53 null mice alone. Furthermore, there is no increase in the incidence of chromosomal instability or abnormal centrosome number. This finding was unexpected for several reasons. First, the immunodepletion of *Drosophila melanogaster* TACC, D-TACC, results in polyploidy[141]. Furthermore, a D-TACC mutant with decreased expression, 10% of wild-type, resulted in severe embryonic defects associated with centrosomal dysfunction; free centrosomes, polyploidy, aneuploidy, nuclear migration defects, and nuclear fusion defects. However, D-TACC is the only TACC family member in *Drosophila melanogaster* so the functional consequences of interfering with D-TACC may not be seen in mammals. Second, in siRNA treated HeLa cells the depletion of TACC3 caused an increase in the mitotic index[139]. The increased mitotic index was a function of the activation of the mitotic checkpoint due to lagging chromosomes at the mitotic plate. The occurrence of lagging chromosomes correlated with the level of TACC3 depletion. Though the primary effect of TACC3 overexpression is predicted to be a detrimental effect on centrosome and mitotic spindle stability alternative mechanisms are possible. Recently, TACC3 was shown to interact with friend of GATA-1 (FOG-1) and prevent its interaction with the globin transcription factor 1 (GATA-1)[142,143]. FOG-1 and GATA-1 are essential for the differentiation of erythroid and megakaryocyte cells. When TACC3 is present at high levels it sequesters FOG-1 in the cytoplasm preventing its localization to the nucleus where it can interact with GATA-1 and mediate differentiation. Therefore, the potential overexpression of TACC3 as a result of t(4;14) may interact with other regulators of differentiation and potentially limit the differentiation of myeloma plasma cells. In this context, it is interesting that an immature/intermediate plasma cell phenotype is common in t(4;14) myeloma cases[85]. Alternatively, TACC3 dysregulation may affect gene regulation by an interaction with chromatin remodeling complexes. TACC3 has been shown to directly bind the histone acetyltransferase (HAT) hGCN5L2 by GST pull-down and can be co-immunoprecipitated by antibodies against pCAF[144]. Thus, a dysregulation of TACC3 may contribute to the oncogenic process associated with t(4;14).

I.5.2 - FGFR3/ACH/CEK2/JTK4/HSFGFR3EX

When t(4;14) was first identified, in 1997, Chesi et al. proposed that *FGFR3* was the target gene [88]. *FGFR3* was the suggested target gene for several reasons. First, at the time the only known, mapped, and characterized gene around the breakpoint region was *FGFR3*. Second, *FGFR3* was only expressed in t(4;14) positive samples. Third, the *FGFR3* paralog, *FGFR1*, was believed to be the target gene of the breast cancer 8p11 amplicon[145], suggesting that dysregulation of FGFR family members could be oncogenic. Fourth, some t(4;14) positive samples expressed *FGFR3* alleles with known activating mutations. Furthermore, these mutations were expressed in a mono-allelic fashion. Moreover, the mutations were somatic mutations based on the fact that similar germline mutations lead to dwarfism and none of the positive individuals were dwarfs.

The *FGFR3* gene was cloned by two different groups within a six month period. First, Keegan et al. identified a cDNA clone containing *FGFR3* during a search for tyrosine kinases expressed in leukocytes[146]. The cDNA clone was identified by screening a cDNA library, generated from the human chronic myelogenous leukemia cell line K562, with a chicken tyrosine kinase probe under low stringency conditions. Second, Thompson et al. cloned a nearly identical cDNA during a search for the Huntington disease gene[147], which was mapped to 4p16[148-150]. The cDNA clone was identified by screening a human fetal brain cDNA library with a genomic probe that mapped within the Huntington disease region and hybridized in cross-species tests, suggesting the presence of exons. The cloning of *FGFR3* from the Huntington disease region mapped the gene to 4p16, but approximately 400 bp of the 5' end, including the start codon, was not cloned by Thompson et al.

Although *FGFR3* did not turn out to be the causative gene of Huntington disease it proved to be a causative gene in another genetic disease. The causative gene of achondroplasia, one subtype of dwarfism, was linked to 4p16 by linkage mapping and subsequently point mutations in *FGFR3*, which result in missense amino acid substitutions, were identified in affected individuals[151-155]. Subsequently, a variety of different *FGFR3* mutations were identified in most cases of human dwarfism (Reviewed in detail by Vajo et al. and Passos-Bueno et al.)[95,156]. The severity of skeletal dysplasia is variable and related to the specific mutation (Table I.3), however, severity ranges from lethal (thanatophoric dysplasia type I and type II), to severe (achondroplasia), or mild (hypochondroplasia). Moreover, other diseases with similar phenotypes are also linked to mutations within the coding region of *FGFR3*, these include severe achondroplasia with developmental delay and acanthosis nigricans (SADDAN), cranisynostoses associated or without other limb malformations (CRS), platyspondylic lethal skeletal dysplasia, San Diego type (PLSD-SD), and Crouzon with acanthosis nigricans (C+AN). In general the mutations in *FGFR3* are activating mutations that lead to increased FGFR3 mediated signaling cascades. The mechanism of FGFR3 activation varies between the different mutations. FGFR3 signaling is normally a regulated process whereby ligand binding induces receptor dimerization, which causes auto- and trans-phosphorylation to initiate the signaling cascade and eventually the downregulation of the signaling cascade by receptor internalization and degradation. The proposed mechanism of action for the mutations is ligand independent receptor activation. However, this is not the case for all mutations. The acondroplasia (ACH) specific G380R mutation does not cause ligand independent activation but does prevent the down regulation of receptor mediated signaling by preventing the internalization of ligand dimerized FGFR3[157]. Some of the mutations do fit the hypothesis of ligand independent receptor activation. Interestingly four different mutations, which replace the wild-type amino acid with a cysteine, exist within six amino acids of the extracellular juxtamembrane region, G370C, S371C, Y373C, and G375C. The first three are associated with the lethal thanatophoric dysplasia type I (TDI) phenotype while G375C is associated with the viable ACH phenotype (Table I.5). All four cause ligand independent receptor dimerization, receptor phosphorylation, MAPK phosphorylation, and target gene transcription[158]. The G370C and S371C mutations, which are exclusively associated with the lethal TDI phenotype, are unresponsive to ligand, suggesting that the mutations induce a receptor confirmation maximizing downstream effects[158]. The Y373C and G375C mutants are still responsive to ligand and require ligand to maximize downstream effects[158]. This dependence on ligand may explain why G375C confers a viable phenotype and why the Y373C mutant confers the unique PLSD-SD phenotype. Moreover, four different mutations exist at position 650 (Table I.3). The K560M (TDI/SADDAN) and K650E thanatophoric dysplasia type II (TDII) result in 18 fold and 9 fold increases in the auto-phosphorylation of FGFR3 compared to wild-type[159]. The K650Q and K650N mutations associated with the mild hypochondroplasia (HCH) phenotype result in 3.7 and 4.9 fold increases in auto-phosphorylation of FGFR3. Therefore, the effect of *FGFR3* mutation is not universal as each mutant is associated with a specific disease phenotype and biochemical effects.

The potential contribution of FGFR3 to a malignant condition was first established by Chesi et al. in multiple myeloma[88], where FGFR3 expression was shown to be associated with t(4;14) and a subset of the FGFR3 expressing cell lines and patients were shown to harbor mutations associated with TDI and TDII. Based on the observations in myeloma and their observation of *FGFR3* expression in bladder and cervix epithilia, Cappellen et al. investigated the expression and mutation status of

FGFR3 in bladder and cervix carcinoma[160]. Greater than 90% of the tumours from either site expressed *FGFR3*. Moreover, 35% of the bladder and 25% of the cervix tumours had *FGFR3* mutations associated with TDI and TDII. This high incidence of presumptively activating FGFR3 mutations in both cancers resulted in a detailed analysis of several large cohorts. It is now widely accepted that approximately 40% of bladder cancers have *FGFR3* mutations[160-164]. Interestingly, these mutations correlate with a lower tumour grade and a low recurrence rate[162-165]. The initially reported incidence of *FGFR3* mutations in cervix carcinoma has not been replicated by two subsequent publications that reported occurrence rates of 1.9% and 3.5%[166,167]. Moreover, Sibley et al. screened a panel of stomach, rectal, colon, prostate, ovarian, breast, brain, and renal tumours and did not find a single *FGFR3* mutation[167].

In multiple myeloma, the expression of *FGFR3* is associated with t(4;14). Initially, the occurrence of *FGFR3* mutations in t(4;14) positive samples was thought to be high as 2/5 (40%) cell lines, and 1/4 (25%) patient samples had characterized mutations[88]. However, in expanded cohorts from multiple groups missense mutations are quite rare with only 6% of t(4;14) positive/FGFR3 positive patients having mutations[88,118,119,168-171]. To date, only 7 different missense mutations have been identified in myeloma patients and cell lines (Table I.3). Furthermore two different in-frame genomic deletions of the proper stop codon have been identified in one patient and one cell line (Table I.3). Several of the identified mutations are known activating mutations as they occur in TDI and TDII. These included the R248C mutation identified in two patients[118,169], the Y373C mutation in the KMS-11 cell line[88,96], the K650M mutation in one patient[88], and the K650E mutation in one patient and the OPM-2 cell line[88,119]. Interestingly, the one myeloma patient expressing the K650M mutation, MM.T1, also expresses the wild-type K650 allele[35,88]. The expressed wild-type allele originates from the chromosome 4 involved in the translocation as the two alleles can be differentiated by two silent polymorphisms. Therefore, the *FGFR3* mutation occurred after the translocation and both sub-clones remain. The remaining mutations have not been identified in dwarfism or related syndromes. Two similar genomic deletions encompassing the proper stop codon were identified in the MM5.1 cell line and one patient[35,170]. Though this type of mutation has not been identified in chondroplasia patients they should be synonymous with the TDI point mutations, which delete the stop codon and result in a protein with an additional 141 C-terminal amino acids, as both deletions maintain the original reading frame. Therefore, these mutations are very likely activating mutations. The remaining mutations are either unique to myeloma; Y241C in one patient[171], and G382D in the KMS-18 cell line; or present in both bladder cancer and myeloma like the F384L mutation present in one patient and the LP-1 cell line [35,164,168]. Interestingly, the F384L mutation was found in 2/100 normal individuals; it may actually reflect a rare but naturally occurring polymorphism and not a tumor-specific somatic mutation [168].

Table I.3 – Known FGFR3 Mutations

Type of Dwarfism	**Mutation**	**Present in Myeloma (Other Neoplasia)**
TDI	R248C **(PLSD-SD)**	Yes, 2 patients [118,169] (Bladder-14 Patients)[160-165]
	S249C **(PLSD-SD)**	(Bladder-122 patients)[160,163-165] (Cervical-4 patients)[160,162,166]
	G370C	(Bladder-7 patients)[160,162,163,165]
	S371C	

	Y373C **(PLSD-SD)**	Yes (KMS-11)[88,96] (Bladder-40 patients)[161-164]
	K650M **(SADDAN)**	Yes, 1 patient***[88] (Bladder- 2 patient)[162,163]
	X807L**	
	X807G** **(PLSD-SD)**	
	X807R**	
	X807C**	
	X807W** **(PLSD-SD)**	
TDII	K650E	Yes, 1 patient (OPM-2)[119 88] (Bladder- 9 patients)[160,162-164]
Achondroplasia	G375C	
	G380R	
Hypochondroplasia	N328I[172]	
	I538V	
	N540T	
	N540K	
	K650Q[159]	(Bladder-1 patient)[161]
	K650N[159]	
Related Disease	P250R **(CRS)**	
	A391E **(C+AN)**	
Not Associated with Chondroplasia	Y241C	Yes, 1 patient[171]
	E322K	(Colorectal- 1 patient)[173]
	G382D*	Yes (KMS-18)[174]
	F384L	Yes, 1 patient (LP-1)[35,168] (Bladder- 8 patients)[164]
	A391E	(Bladder-1 patient)[165]
	Δ795-808**	Yes (MM5.1/MM5.2)[35]
	Δ797-811**	Yes, 1 patient[170]

Unless otherwise noted all data is from Passos-Bueno et al. and references therein[95].

*Generally noted as G384D since the IIIb isoform is expressed primarily in KMS-18

** Results in a deletion of the stop codon. All deletions and point mutations remain in the same reading frame and an additional 423 bp (141 amino acids) are translated until the next in-frame stop codon is reached.

***This patient expressed both the mutated K650M isoform and the wild-type isoform K650K. This wild-type allele did not originate from the non-involved allele as two polymorphic markers allowed the specific translocated allele to be recognized.

The expression of *FGFR3* was proposed to be the transforming event of t(4;14) positive myeloma and thus a number of groups have tested the transforming capability of *FGFR3*. Initial work by Webster et al. suggested *FGFR3* was not transforming[175,176]. Neither, wild-type nor the K650E mutant forms of *FGFR3* caused transformation of NIH 3T3 cells even though K650E induces ligand independent phosphorylation of the receptor. However, if the extracellular and transmembrane regions were deleted the K650E mutant could transform NIH 3T3 cells when targeted to the plasma membrane. These initial observations were not replicated in subsequent studies. Chesi et al. found the K650E (OPM-2), Y373C (KMS-11), and Δ795-808 (MM5.1) mutants but not wild-type or F384L (LP-1) mutants induced transformation of NIH 3T3 cells[35]. The discrepancy originates from the use of different promoters, cytomegalovirus (CMV) (Webster) and elongation factor 1-alpha (EF-1α) (Chesi), and the higher level of expression attained with the EF-1α promoter. As predicted, Chesi et al. did not observe transformation when a CMV promoter was used to drive *FGFR3* expression. The inability of the F384L mutant to transform NIH 3T3 cells and its detection in 2/100 normal individuals supports the suggestion of this mutant being a natural polymorphism and not an acquired mutation[168]. Moreover, the mechanism of transformation is through the RAS/RAF/MEK/MAPK pathway as K650E mediated transformation can be inhibited by RAF and RAS dominant negative constructs with increasing effectiveness, respectively. Similarly, Ronchetti et al. showed K650E (OPM-2) and Y373C (KMS-11) but not G382D (KMS-18) or wild-type *FGFR3* could transform NIH 3T3 cells[174]. The expression of *FGFR3* in the transformed NIH 3T3 cells by the CMV2 promoter was significantly greater than the expression levels seen in the t(4;14) positive cell lines with the same mutations. Therefore, the ability of K650E and Y373C to transform NIH 3T3 cells is dependent on the expression level attained. Although the G382D (KMS-18) mutant did not transform NIH 3T3 cells the receptor is still functional. In the absence of ligand the FGFR3 mutants expressed by OPM-2 and KMS-11 are constitutively phosphorylated while the G382D mutant in KMS-18 is not phosphorylated, however, in the presence of ligand the G382D FGFR3 mutant is phosphorylated, suggesting the mutant receptor is functional. In all three cell lines MAPK is phosphorylated in the absence of ligand but in the presence of ligand the amount of phosphorylated MAPK is increased as predicted by Chesi et al[35]. The constitutive phosphorylation of the Y373C mutant in the absence of ligand is maintained when expressed in the t(4;14) negative cell line U266 and identical to the observations in KMS-11, ligand increases the amount of phosphorylated receptor and MAPK.

The study of Plowright et al. tested the effects of wild-type and K650E mutant *FGFR3* on the IL-6 dependent murine plasmacytoma cell line, B9[130]. Expression of the K650E mutant in B9 cells resulted in IL-6 independent proliferation and the addition of IL-6 or ligand, augmented proliferation. The wild-type receptor did not result in IL-6 independent growth nor did the addition of IL-6 increase the proliferation rate compared to controls. However, the addition of ligand to B9 cells expressing wild-type FGFR3 increased the proliferation of B9 cells. Unlike the K650E expressing mutants, only sub-clones expressing high levels of the wild-type receptor could achieve IL-6 independence. Therefore, the contribution of wild-type *FGFR3* in t(4;14) positive patients will likely be related to the expression level, which is quite variable between patients[100,121,122]. The study by Li et al. tested the *in vivo* transforming capability of human wild-type and K650E mutant *FGFR3* in BALB/c mice[129]. Bone marrow cells were isolated from donor BALB/c mice, transfected with a retroviral construct containing the wild-type or K650E mutant receptor, and transplanted in irradiated recipient BALB/c mice. The mice receiving bone marrow (BM) transfected with the K650E mutant developed lethal leukemia-like disease within 4-6 weeks with mean WBC counts of approximately 100 000/μl (~10 000/μl in control mice). The mice receiving BM transfected with wild-type *FGFR3* became morbid around 1 year, while control mice survived beyond 18 months. The wild-type *FGFR3* expressing mice had elevated WBC counts (~25 000/μl) and infiltration of major organs with lymphoid cells.

Although control mice survive, suggesting a latent oncogenic effect for overexpressed wild-type human *FGFR3* in BALB/c mice[129], the potential mechanism is unclear. Unfortunately, controls were not included to verify that the expression of human FGFR3 in BALB/c mice did not result in a chronic B- or T-cell stimulation which caused the proposed acute lymphocytic leukemia (ALL) like death of these mice. Furthermore, recent work with transgenic mice expressing activated forms of *FGFR3* in isotype switched B cells has not resulted in gammopathy or myeloma/leukemia like disease, even after 2 years of follow-up[177]. Moreover, the MM.T1 patient identified by Chesi et al. expresses both the wild-type and K650M activating form of *FGFR3* at near equal levels[35,88], suggesting the acquired mutation does not provide a substantial growth advantage, since it does not predominate. However, serial analysis of this patient may have identified tumour selection, since Sibley et al. found a patient expressing wild-type *FGFR3* at diagnosis but at relapse the K650E mutation was detected[119]. Whether this reflects an acquired mutation during treatment or the outgrowth of a rare sub-clone present at diagnosis was not determined. Finally, individuals with ACH and HCH do not have increased incidences of malignancy, although these syndromes are characterized by different mutations. Therefore, the actual effect of *FGFR3* expression and mutation in t(4;14) myeloma remains elusive.

I.5.3 - LETM1

Due to the proximity of *LETM1* to known t(4;14) breakpoints on chromosome 4 we proposed that it may be a potential target gene[1]. A sequence analysis comparison between murine and human *LETM1* identified several evolutionarily conserved domains. These include; an N-terminal type II (N_{in}-C_{out}) transmembrane domain, several coiled-coil domains; putative protein kinase C (PKC), casein kinase 2 (CK2), and tyrosine kinase phosphorylation sites, a C-terminal leucine zipper motif and two EF-hand motifs. The second EF-hand motif is perfectly identical to the consensus EF-hand Ca^{2+} binding site, while the first EF-hand diverges at three positions and is unlikely to bind calcium. Therefore, LETM1 was characterized as a novel EF-hand Ca^{2+} binding protein with a transmembrane domain that may be regulated by calcium levels. Recent work showed LETM1 and the yeast homolog YOL026 localize to mitochondria by a domain present in the N-terminal 167 amino acids upstream of the PKC/CK2 phosphorylation site and transmembrane domain[178,179]. The yeast homolog localizes to the inner mitochondrial membrane and when deleted results in mitochondrial dysfunction that can be complemented by the human *LETM1* gene[179].

I.5.4 - WHSC1/MMSET/NSD2/TRX5

MMSET was the second proposed t(4;14) target gene based on several factors [91]. First, *MMSET* transcripts were overexpressed in t(4;14) positive samples[91]. Second, several conserved protein domains in MMSET, particularly the SET domain, are also found in proteins involved in chromatin remodeling and epigenetic gene regulation. Particularly, *MMSET* is very similar to the *HRX/MLL* gene that is involved in plethora of different translocations in AML and ALL. Furthermore, the *MMSET* paralogs, *NSD1* and *WHSC1L1/NSD3*, were identified as partner genes of *NUP98* translocations in AML[180-184].

The *WHSC1* gene was cloned by two independent groups within a three month period in 1998. The first indication that a gene existed at this genetic locus was made by Pribill et al. who were searching for expressed genes within the 2 Mb Huntington disease region and identified a gene similar to *ASH1* from *Drosophilia melanogaster*[185,186]. However, this study identified a large number of genes and did not attempt to characterize each gene. The first characterization of *WHSC1* was made by Stec et al. during a search for genes within 165 kb WHS critical region[90]. The WHS critical region was previously identified by mapping overlapping deletions of 4p16 found in WHS patients[187]. Stec et al. mapped a number of expressed sequence tags (ESTs) to this region by comparative sequence analysis and the translation product of one of the many ESTs was similar to the *Drosophilia melanogaster* protein ASH1[185]. Since, WHS and ASH1 mutants have phenotypic malformations the gene encoding this EST was a promising WHS candidate gene. Subsequently, they were able to characterize the entire open reading frame (ORF) and multiple splice variants using overlapping EST sequences, exon prediction programs and RT-PCR. Almost simultaneously Chesi et al. identified *WHSC1* during a search for genes interrupted by t(4;14) translocations, however, they called the gene *MMSET*[91]. Using the GRAIL2 exon prediction program they identified exon 3 of *MMSET* within cosmid L184d6, which contains a number of t(4;14) breakpoints. To confirm that this region was expressed exon 3 was amplified and used to probe a Northern blot. Subsequently, a testis cDNA library was screened with the exon 3 probe and this identified the two principal alternatively spliced ORFs encoding MMSET I and MMSET II.

A number of different transcripts originate from the *MMSET* locus. Nine different transcript variants are listed in the Entrez Gene database although other unlisted variants exist. The majority of these transcript variants differ in the alternatively spliced 5' UTR. Ultimately, only four principal protein variants are encoded by the different transcripts; MMSET I, MMSET II, MMSET III, and RE-IIBP (Figure I.5)[90,91,122]. Transcripts encoding the MMSET protein variants typically initiate in *MMSET* exon 1a, exon 1, or exon 2a (*WHSC1* exon 1)[90,91]. The majority of these transcripts will splice directly to *MMSET* exon 3 (*WHSC1* exon 4) that contains the proper translation initiation site though some included various combinations of exons 2a, 2b, 2c, 2d, and 2e. Downstream of *MMSET* exon 3 two different alternative splicing events produce the different MMSET protein products. First, MMSET III is created by alternative splicing of exon 4 to exon

4a instead of exon 5 as an in-frame stop codon is present in exon 4a[101,122]. Second, alternative splicing of *MMSET* exon 10 to exon 11 or exon 12 results in either MMSET I or MMSET II, respectively. The MMSET I protein is created as an in-frame stop codon is present in exon 11. However, if *MMSET* exon 10 is alternatively spliced to exon 12, translation continues to the stop codon in exon 24 producing the full length protein, MMSET II. The remaining protein product, RE-IIBP, is produced from an alternative transcription event initiating within *MMSET* intron 9 which follows the identical splicing patterns as the MMSET II encoding transcripts [188]. The first translation initiation site in this transcript in-frame with the MMSET II reading frame is located in *MMSET* exon 15 and translation from this point produces RE-IIBP.

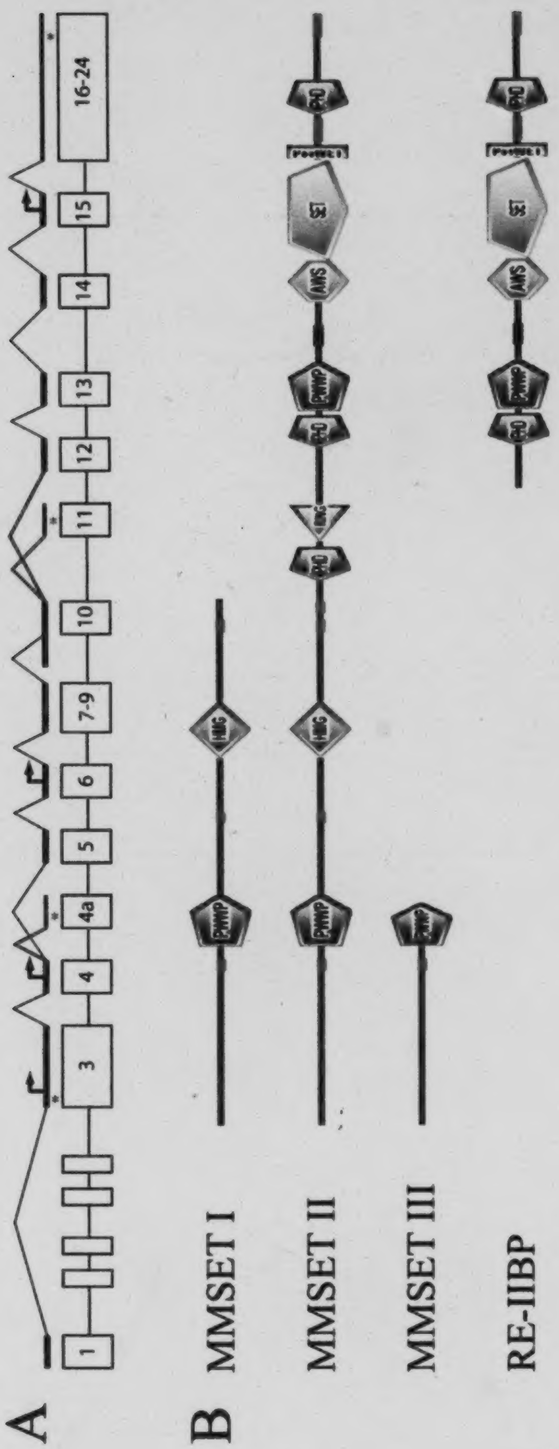

A
1
3
4
4a
5
6
7-9
10
11
12
13
14
15
16-24
B
MMSET I
MMSET II
MMSET III
RE-IIBP
PWWP
HMG
PHD
AWS
SET
PostSET

Figure I.5 – Alternative Splicing and Initiation of Transcription of MMSET Results in Various Wild-type Proteins

A) The rough exon-intron structure of the *MMSET* gene is shown. Thin black lines indicate the proper splicing pattern producing MMSET II encoding transcripts. Alternative splicing events producing MMSET III and MMSET I encoding transcripts are indicated by thick red lines. In-frame stop codons are indicated by a red asterisks. The proper MMSET translation initiation site in exon 3 is indicated by a green arrow, while the proposed alternative translation initiation sites in exons 4 and 6 identified by Chesi et al. are indicated by blue arrows[91]. The point of transcription initiation for RE-IIBP encoding transcripts is indicated by a green line and the translation initiation site in exon 15 is indicated by a green arrow. B) Scale diagrams of the MMSET variants and the conserved protein domains identified by the S.M.A.R.T. protein prediction program[189,190].

Very little is known about the function of *WHSC1/MMSET*. No functional analysis of *WHSC1* exists in the literature. In the myeloma field, several expression studies exist but functional information is extremely limited. In general it is accepted that the expression of *MMSET* is elevated in patients with t(4;14)[82,91,117]. Moreover, some investigators have found similar expression levels in rare t(4;14) negative patients[100,112,117]. The only published functional information on *MMSET* was a comment in the discussion of a paper by Chesi et al. and no supporting data was shown or subsequently published[88]. They commented that MMSET I inhibits transformation of NIH 3T3 cells when co-transfected with known oncogenes, suggesting this N-terminal variant may act as a tumour suppressor, and that the full length MMSET II is not transforming. The only other functional comment is from an abstract presented at the 2004 annual meeting of the American Society of Hematology claiming that transgenic expression of *MMSET* was not detectable in normal B cells, however, B cell tumours expressing *MMSET* occurred in these mice, suggesting *MMSET* dysregulation is transforming[191].

The conserved protein domains present in *MMSET* can be used to identify conserved orthologs and paralogs. *WHSC1/MMSET/NSD2* has two human paralogs, *WHSC1L1/NSD3* and *NSD1* which arose from *en bloc* chromosomal segment duplications (The NSD gene family). These paralogs are characterized by a similar protein structure of N-terminal to C-terminal protein domains; PWWP, PHD, PHD/RING, PHD, PWWP, AWS, SET, ps, PHD (Figure I.5). In particular the PWWP and SET domains link this gene family to a large number of proteins providing some insight into the potential function of *MMSET*[192,193].

The most important domain encoded by *MMSET* is believed to be the SET domain. The SET/tromo domain is a conserved protein domain initially identified in the *Drosophila melanogaster* proteins Su(var)3-9, Enhancer of zeste, and Trithorax[194,195]. The Su(var)3-9 gene is a suppressor of position-effect variegation, which is the regulation of gene expression by proximity to heterochromatin. Furthermore, the Polycomb group (Enhanzer of zeste) and the Trithorax group (Trithorax) of proteins negatively and positively regulate the expression of developmentally regulated genes, respectively. Since all three genes regulate gene expression the shared SET domain was proposed to be the protein domain mediating gene regulation. We now know SET domains regulate gene expression by methylating nucleosomal histone tails resulting in permissive or repressive chromatin states. The change in chromatin states involves complex post-translational modifications (methylation, acetylation, phosphorylation) of the histone tails. Furthermore, as suggested by the different functions of Polycomb and Trithorax group proteins, different methylation events are associated with different chromatin states. Therefore, sequence differences within SET domains likely reflect target specificity (eg. H3-K4 versus H3-K9) and which specific methylation events are catalyzed (eg. mono, di, or tri-methyl).

Sequence identity within the SET domain of different SET domain containing proteins identified four evolutionary conserved subgroups present in yeast, fruit flies, and humans[193]. This analysis placed the human NSD protein family, *ASH1L*, and *HYPB* in a subgroup characterized by the *Saccharomyces cerevisiae* SET2 protein (Figure I.7). The *ASH1L* gene is the human ortholog of the *Drosophila melanogaster* ASH1 gene. The sub-grouping of ASH1 and *MMSET* into the SET2 subgroup is interesting since Stec et al. initially cloned *WHSC1* based on the similarity with *Drosophila melanogaster* ASH1[90]. Recently, the *Drosophila melanogaster* Mes-4/CG4976 and CG1716 genes were identified as the orthologs of *WHSC1/WHSC1L1/NSD1* and *HYPB*, respectively[196]. It is interesting to note that SET2 diverged into at least 3 different genes in Drosophila melanogaster but only the Mes-4 gene diverged further into *WHSC1/MMSET/NSD2*, *NSD1*, and *WHSC1L1/NSD3* in humans (Figure I.6).

The founding member of the subgroup, SET2, is required for the methylation of H3-K36 but not H3-K4 (mediated by SET1) in *Saccharomyces cerevisiae* [197]. Moreover, the *Drosophila melanogaster* ortholog, ASH1, methylates H3-K4, H3-K9, and H4-K20 *in vitro*[198]. However, *in vivo* ASH1 is essential for H3-K4 methylation, but does not methylate H4-K20 and has only limited specificity for H3-K9[199]. Therefore, assuming sequence similarity within the SET domain confers target specificity the H3-K4, H3-K9, H3-K36, or H3-K20 residues are the most likely targets of the NSD proteins. Following this assumption Rayasam et al. showed murine Nsd1 is a functional HMTase with *in vitro* specificity for H3-K36 and H4-K20[200]. Therefore, should the SET domain of MMSET be a functional HMTase the most likely target residues are H3-K36 and H4-K20, but other residues are also potential targets. Furthermore, most studies to date used only antibodies raised against di-methylated residues; however, the possibility exists that a specific histone methyltransferase (HMT) may only catalyze mono or tri-methylation events.

The presence of a PWWP domain is a feature unique to the NSD family and other evolutionarily related *Drosophila melanogaster* Mes-4 orthologs within the SET2 family of HMTases. To date, the only proteins characterized in the S.M.A.R.T database with both PWWP and SET domains are Mes-4 orthologs (WHSC1, NSD1, and WHSC1L1). Therefore, the PWWP domain likely plays a fundamental role in the function of Mes-4 orthologs. The PWWP domain was characterized by Stec et al. after they cloned *WHSC1* and proposed to mediate protein-protein interactions[90,192]. Subsequently, the PWWP domain, was identified in a large number of nuclear proteins including; mutS homolog 6 (MSH6), hepatoma-derived growth factor (HDGF), DNA (cytosine-5-)-methyltransferase 3 alpha (DNMT3a), DNA (cytosine-5-)-methyltransferase 3 beta (DNMT3b), and others[192]. The significance of the PWWP domain is underscored by several missense mutations within the domain which result in human diseases. These include, the S144I MSH6 mutation (HGMD Accession #CM992929) associated with both microsatellite instability low (MSI-L) and microsatellite instability high (MSI-H) HNPCC tumours[201,202], the S270P[v] DNMT3b mutation (HGMD Accession #CM021982) associated with ICF syndrome[203], and most importantly the G1792V and V1796F NSD1 mutations (HMGD Accession #CM030077 and CM032976) associated with Sotos syndrome[204]. The molecular function of the PWWP domain is unclear but several recent reports strongly suggest the domain may be a DNA or chromatin binding domain. First, Qui et al. solved the crystal structure of the murine Dnmt3b PWWP domain and identified a characteristic surface feature that may bind approximately 12 bp of DNA[205]. Deletion of the PWWP domain decreased the *in vitro* DNA binding efficiency of Dnmt3b. However, deletion of the PWWP domain does not affect the *in vitro* methyl transfer activity of Dnmt3b. Second, Ge et al. showed the PWWP domain is sufficient to target Dnmt3a and Dnmt3b to metaphase chromosomes[206]. During interphase the Dnmt3a and Dnmt3b variants localize to heterochromatin *in vivo*. This localization is dependent on the PWWP domain, but the PWWP domain is not sufficient for this localization. The mouse equivalent of the S270P DNMT3b mutation completely abrogates the localization of Dnmt3a and Dnmt3b variants to metaphase chromosome and interphase heterochromatin. Third, Chen et al. showed the PWWP domain is required for the localization of Dnmt3a and Dnmt3b variants to pericentric heterochromatin and the methylation of major satellite repeats[207]. Furthermore, they showed that Dnmt3a did not bind DNA *in vitro* and the previously characterized *in vitro* DNA binding of Dnmt3b was unspecific. Similarly, the PWWP domain of SPBC215.07c in *Saccharomyces pombe* does not bind DNA[208]. The unspecific DNA binding capability of Dnmt3b is likely due to a large number of unconserved basic amino acids within this PWWP domain[192,208].

---

[v] In the discovery paper by Shirohzu et al. the amino acid numbering was based on NCBI Accession #AAF04015 and not NP_008823 which is the reference sequence used by the HGMD.

Based on the available evidence the PWWP domain is likely involved in targeting proteins to chromatin, but the exact interactions remain to be determined. Furthermore, disease associated missense mutations highlight the absolute importance of this domain. Moreover, in the case of the *MMSET* paralog, *NSD1*, the mutations in the C-terminal PWWP domain suggests that either both PWWP domains are required for proper function or only the C-terminal PWWP domain is required.

Although the SET and PWWP domains of *MMSET* are the only universally conserved protein domains in all Mes-4 orthologs, several other protein domains are encoded by *MMSET*. These include an N-terminal HMG domain, three C-terminal PHD fingers, and a RING domain. The HMG box is unique to *MMSET* as neither paralog, *NSD1* nor *WHSC1L1*, contain this domain (Figure I.6). The HMG domain is a conserved DNA binding domain present in numerous proteins[209]. Since *MMSET* proteins contain only a single HMG box this domain likely provides a sequence specific DNA binding activity as proteins with single or multiple HMG box domains typically bind DNA in a sequence specific or unspecific manner, respectively. The PHD and RING domains identified in MMSET present several complexities. First, depending on the prediction program used the number of PHD domains and the presence or absence of a RING domain varies (SMART, 3 PHD and 1 RING; PFAM, 3 PHD; Prosite, 2 PHD and 1 RING; CDART, 1 PHD). The PHD and RING domains are both zinc finger sub-families with a classical C4HC3 and C3HC4 sequence motif, respectively. The PHD fingers are conserved in the *Drosophila melanogaster* Mes-4 ortholog and the human paralogs *WHSC1L1*, and *NSD1*. In general, PHD fingers, are thought to mediate protein-protein interactions[210]. The PHD fingers of ATP-dependent chromatin assembly factor (ACF1), part of the chromatin remodeling complex, bind to core regions of histones H2A, H2B, H3, and H4 in a zinc dependent fashion[211]. Similarly, the ability of a protein segment from p300, consisting of the BROMO and PHD domains, to bind nucleosomes in a nucleosome retention assay was dependent on the PHD domain, and the PHD domain alone could bind nucleosomes in an electrophoretic mobility shift (EMSA) assay[212]. However, this activity is dependent on the PHD domain from p300 as replacement with a PHD domain from the *MMSET* relative, *MLL*, prevents nucleosome retention. In contrast, Gozani et al. showed the ING2 PHD domain binds PtdIns(5)P and PtdIns(3)P in a zinc dependent fashion[213]. Furthermore, they showed the ACF1 PHD domain also bound PtdIns(5)P. These variable results highlight the problem protein domain prediction programs have differentiating highly similar domains. This is an established problem between the PHD and RING zinc finger subfamilies as variant motifs exist and the spacing between cysteine and histidine residues are variable and overlapping in both sub-families[214-217]. Furthermore the results of Gozani et al. suggest similar problems exist between PHD fingers and the characterized lipid binding zinc finger FYVE sub-family with a C4 consensus sequence[213,218]. The predicted RING domain is unique to *MMSET* as both *NSD1* and *WHSC1L1* along with the *Drosophila melanogaster* ortholog Mes-4 do not contain RING domains. However, this may represent problems with the protein prediction programs as overlapping PHD and RING domains are identified with the SMART prediction program in all Mes-4 orthologs, but *MMSET* is the only one where the RING domain has the higher prediction value. Therefore, should the prediction of a RING domain be correct, MMSET II may have the characteristic E3 ubiquitin ligase activity of a RING domain[218,219]. Ultimately, the function of the HMG, PHD, and RING domains in MMSET will need to be determined. However, a number of missense mutations within the various PHD domains of NSD1 result in Sotos syndrome[vi][204,220-222]. Therefore, each individual PHD domain is essential to the function of *NSD1*, and likely by extrapolation, they are also essential for the function of *MMSET*.

The potential significance of *MMSET* dysregulation can be predicted based on our current understanding of the different protein domains within the MMSET protein variants. Most likely MMSET acts as an epigenetic regulator of gene expression. The potential mechanism of action would involve the targeting of MMSET to chromatin by the PWWP and PHD domains. Targeting specificity would be mediated by the HMG box with additional targeting specificity or affinity potentially provided by the PWWP domains. After or before MMSET is targeted to chromatin the PHD domains may mediate important protein-protein interactions essential for the formation of a functional MMSET complex. At this point the potential RING domain may target chromatin or co-factors for ubiquitinylation to potentially create a competent chromatin state for epigenetic change or maintenance. Finally, the HMTase activity mediated by the SET domain likely mediates the transfer methyl groups to H3-K36 or H4-K20, which confers an epigenetic mark on the targeted chromatin domain resulting in gene regulation. The dysregulation of this potential activity could alter gene expression profiles at a global level and potentially result in transformation or the selection of cells with transforming potential.

[vi] HGMD Accession numbers CM030073, CM030074, CM030075, CM031302, CM033420, CM032974, CM030262 and the recently identified R2152Q mutation by Melchior et al.[220]

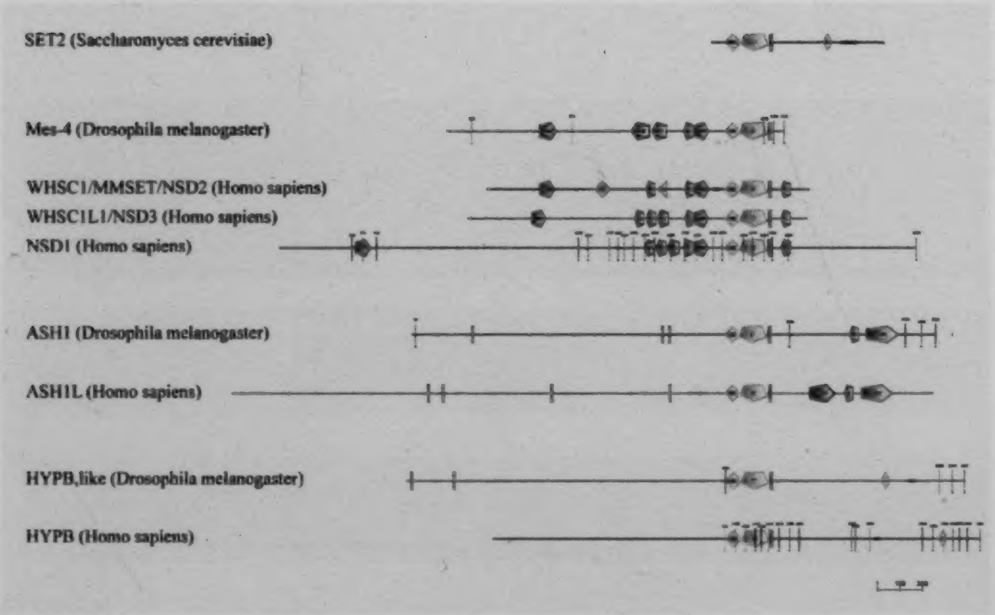

SET2 (Saccharomyces cerevisiae)
Mes-4 (Drosophila melanogaster)
WHSC1/MMSET/NSD2 (Homo sapiens)
WHSC1L1/NSD3 (Homo sapiens)
NSD1 (Homo sapiens)
ASH1 (Drosophila melanogaster)
ASH1L (Homo sapiens)
HYPB,like (Drosophila melanogaster)
HYPB (Homo sapiens)

Figure I.6 – The SET2 Histone Methyl-Transferase Family

Based on sequence similarity within the SET domain the human *WHSC1*, *WHSC1L1*, *NSD1*, *ASH1L*, and *HYPB* genes were assigned to the *Saccharomyces cerevisiae* SET2 gene family[193]. Three subgroups are defined by the conserved protein domains and the order of these domains in the *Drosophila melanogaster* orthologs. The Mes-4 subgroup includes three human orthologs; *WHSC1/MMSET/NSD2*, *WHSC1L1/NSD3*, and *NSD1*. The ASH1 and HYPB, like subgroups only contain a single human ortholog, *ASH1L* and *HYPB*, respectively. Primary protein sequence was analyzed with the S.M.A.R.T. web-tool (http://smart.embl-heidelberg.de) to identify conserved protein domains and to generate scale diagrams[189,190]. Sequences used in the analysis were retrieved from the Entrez sequence database using the following accession numbers; SET2 (NP_012367), Mes-4/CG4976-PA (NP_733239), WHSC1 (NP_579890), WHSC1L1 (NP_075447), NSD1 (NP_071900), ASH1/CG8887-PA (NP_524160), ASH1L (NP_060959), HYPB, like/CG1716-PA (NP_572888), HYPB (NP_054878).

I.5.5 - WHSC2/NELF-A

We proposed that *WHSC2* is a potential target gene of t(4;14) due to its proximity to known chromosome 4 breakpoints[1]. Similar to *WHSC1*, the *WHSC2* gene was discovered by a BLAST analysis of the 165 Kb WHS critical region against dbEST[187]. Characterization of one EST led to the cloning and characterization of the human *WHSC2* and mouse *Whsc2* genes[223]. Transcription from the human gene produces a uniform transcript of 2.4 Kb expressed in all adult and fetal tissues. The transcript contains a 1584 kb ORF encoding a predicted 57 kDa protein composed of 528 amino acids. The protein is predicted to contain an N-terminal transmembrane domain, two putative nuclear localization signals, and a short coil-coiled domain. Based on these features WHSC2 was predicted to localize to the nuclear membrane. Almost simultaneously *WHSC2* was characterized by another group looking for genes regulated by HIV Tat in endothelial cells[224]. Addition of HIV Tat to endothelial cultures resulted in substantially decreased levels of *WHSC2* transcripts. Similar to the observations of Wright et al. *WHSC2* transcripts were detected in all human tissues and cell lines tested. Using cell fractionation and transient transfection with green fluorescent protein (GFP) tagged WHSC2 they confirmed the predicted nuclear localization. Interestingly, this group did not identify a transmembrane domain but did identify a helix-loop-helix domain and thus proposed WHSC2 may act as a transcription factor. This hypothesis was confirmed when NELF-A, a subunit of the NELF complex that represses RNA polymerase II elongation[225-227], was cloned and shown to be identical to *WHSC2*[228]. Therefore, WHSC2 is a critical component in the repression of RNA polymerase II elongation. Based on this working model the potential dysregulation of *WHSC2* by t(4;14) could have dramatic effects on the regulation of transcription at a global level, and thus is an attractive candidate t(4;14) target gene.

Materials and Methods

MM.1 – Origin of Analyzed Samples

All of the samples studied in this body of work are of human origin. This includes a series of cell lines collected from various groups around the world and a large cohort of patient samples collected by our working group.

MM.1.1 - Cell Lines

The myeloma cell lines RPMI 8226, U266, and S6B45 were generously supplied by S. Treon, Dana-Farber Cancer Institute. The myeloma cell lines KMS-11, KMS-12BM, KMS-12PE, KMS-18, KMS-26, KMS-28BM, KMS-28PE, KMS-34 were generous gifts from T. Otsuki, Kawasaki Medical School. The myeloma cell line JIM3 was a kind gift from L. Bergsagel, Mayo Clinic (Scottsdale). The Burkitt's lymphoma cell lines Daudi and Raji were generously supplied by A. Shaw, Cross Cancer Institute. The ovarian cancer fibroblast cell line HeLa was kindly provided by M. Hendzel, Cross Cancer Institute. The myeloma cell line NCI-H929 was purchased from American Type Culture Collection (Manassas, VA). The OPM-2 and LP-1 myeloma cell lines were purchased from the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). All cell lines were maintained in recommended media at $37^{O}C$ in 5% CO_2.

MM.1.2 – Patient Samples

The study was approved by the University of Alberta/Capital Health Authority and Alberta Cancer Board research ethics boards. All patient samples were collected after informed consent at the Cross Cancer Institute, University of Alberta Hospital, or Canadian Blood Services (formerly the Canadian Red Cross) and all three reside within the University of Alberta campus. The analyzed patient tissues originated from bone marrow aspirations, venous peripheral blood phlebotomy, and peripheral blood apheresis products, both fresh and cryopreserved, from patients undergoing stem cell mobilization procedures. We generally receive 30 ml of venous peripheral blood drawn into sodium heparin tubes. The volume of bone marrow aspirate and mobilized blood varies from 1-5 ml. A total of 589 BM, 142 PBMC, and 5 mobilized blood (MB) samples collected between October 1994 and April 2004 were analyzed from 473 individuals.

MM.1.3 – Purification of Bulk Samples

The purification of all patient samples was performed in the laboratory of Dr. Linda Pilarski at the Cross Cancer Institute. All samples were stored at room temperature after clinical isolation and the majority of samples entered the sample purification process within 1 hour. Single cell suspensions of bone marrow aspirates were generated by forcing the aspirate through a wire mesh. The single cell bone marrow suspensions and peripheral blood samples were diluted with 1x PBS and separated into equal 25 ml aliquots in 50 ml conical centrifuge tubes. The diluted samples were underlayed with 15 ml of Ficoll-Paque^{TM} Plus (Amersham Pharmacia Biotech, Uppsala Sweden) and processed according to the manufacturer's protocol. The mononuclear cells isolated from the density gradient were aliquoted for long-term storage in 1.5 ml microcentrifuge tubes. For downstream RNA work 2-10 x 10^6 cells were suspended in 1 ml of TRIzol Reagent (Invitrogen, Carlsbad CA) and stored at -80^{O}C. Alternatively, dry cell pellets consisting of 1-10 x 10^6 cells were stored at -80^{O}C for future DNA or protein based experiments. With an average sample purification time of 1 hour the majority of samples were aliquoted and stored within 2 hours of clinical isolation.

The RNA and DNA samples from the cell lines were isolated from cultures in the log phase of their growth curve. The cells were washed twice with 1x PBS and then stored for subsequent experimental usage as outlined for the patient samples.

MM.1.4 – Plasma Cell Purification

Plasma cells were purified from freshly processed BMMC. The majority were purified on an Automacs Magnetic Cell Separator (Miltenyi Biotec, Auburn, CA) using anti-CD138 microbeads (Miltenyi). Alternatively, plasma cells with a $CD38^{hi}$ and $CD138^{+}$ phenotype were sorted on an Epics® Altra™ Flow Cytometer (Beckman-Coulter, Fullerton, CA). The purity of all sorted samples was verified to be greater than 90% by morphological examination.

MM.2.1 – Clinical Data Extraction

All patient data was extracted from each patients individual paper or electronic chart maintained at the Cross Cancer Institute or University of Alberta Hospital. The data for the initial cohort of patients collected between October 1994 and December 2001 (Keats et al. 2003) was extracted by Dr. Tony Reiman in a blinded fashion during the spring of 2002. Information on diagnosis, patient demographics, baseline staging and clinical features, treatment, response to therapy, progression and survival were collected. All patient records from the two treating institutions were reviewed retrospectively to verify the diagnosis of MGUS and MM. The primary physicians involved in patient management, Dr. Andrew Belch, Dr. Michael Mant, and Dr. Lorree Larratt were unaware of the t(4;14) status of the patients. Subsequently, the clinical data for the additional patients included in the expanded cohort (Keats et al. 2005) and updates of the survival information for patients within the original cohort were extracted by myself, with full knowledge of the t(4;14) status for each patient. The clinical diagnosis of the additional patients was performed retrospectively based on the extracted data with the help of Dr. Tony Reiman. In all instances the diagnosis was determined based on standard criteria [229,230]

MM.2.2 – Clinical Statistics

In the initial cohort the study was designed to look at the prevalence of t(4;14) and its impact on overall survival in myeloma. Secondary analyses included correlation with baseline clinical features, response to therapy, and progression-free survival. Data were analyzed using SAS version 8 for Windows (SAS Inc., Cary, NC) and GraphPad Prism version 3.02 for Windows (GraphPad software, San Diego CA). Categorical variables were compared between two groups using Fisher's exact test. Continuous variables were compared using Student's t-test or the Wilcoxon rank sum test as appropriate. Survival distributions were determined using the Kaplan Meier method and compared using the log rank test. Statistical significance was set at a p-value of 0.05 using two-sided analysis. A Cox regression model was used to adjust for known prognostic factors. Autologous stem cell transplantation was included in the model using landmark analysis. In the expanded cohort the study was designed to determine the clinical impact of *FGFR3* expression and t(4;14) breakpoint type. Survival distributions were determined and compared as noted above.

MM.3 – Molecular Sample Isolation and Analysis

MM.3.1 - RNA and DNA Isolation

Total RNA was extracted from BMMC, PBMC, or purified plasma cells suspended in TRIzol Reagent following the manufacturer's suggested protocol. High molecular weight genomic DNA was isolated from the remaining organic TRIzol Reagent layer, following RNA isolation from the aqueous layer, following the manufacturer's suggested protocol.

MM.3.2 – Synthesis of cDNA for RT-PCR Applications

For standard RT-PCR applications poly-A-tailed RNA was reverse transcribed for 1hr at 42^{O}C followed by enzyme inactivation at 99^{O}C for 3 minutes from 1 μg of total RNA with 500 μM oligo dT15, 500 μM of each dNTP, 10 mM DTT, 50 mM Tris-HCl, 75 mM KCl, 3mM $MgCl_2$, and 200 U Superscript^{TM} RNase H^{-} Reverse Transcriptase (Invitrogen) in a 20 ul reaction.

For quantitative RT-PCR applications 1 μg of total RNA was first treated with 1 Unit of DNase I (Sigma-Aldrich, St. Louis, MO) for 15 minutes and then converted to cDNA with random hexamers using the TaqMan® Reverse Transcription Reagents Kit (Applied Biosystems, Foster City, CA) following the manufacturer's suggested protocol.

MM.3.3 – Standard RT-PCR Reactions

All RT-PCR reactions contained 20 mM Tris-HCl, 50 mM KCl, 2.0 mM $MgCl_2$, 200 μM of each dNTP, 1.0 U Platinum Taq DNA polymerase (Invitrogen, Carlsbad CA) and 0.2 μM of each PCR primer. All primer sequences are listed in Table MM.1.

Table MM.1 - Standard RT-PCR Primers

Primer	Sequence (5'-3')
5' β2-M	CCAGCAGAGAATGGAAAGTC
3' β2-M	GATGCTGCTTACATGTCTCG
JH	ACCACGGTCACCGTCTCCTCA
Iμ1	AGCCCTTGTTAATGGACTTG
Iμ2	CTTTGCAAGGCTCGCAGTGAC
ms6r	CCTCAATTTCCCTGAAATTGGTT
ms5r	AAGAACTGTACGTGATACTG
5' FGFR3-A	GCGCTAACACCACCGACAAG
3' FGFR3-A	CTCCCCTGAGGACAGCCTTGCGAT
5' FGFR3-B	ATGAAGATCGCAGACTTCGGG
5' FGFR3-B	GTAGACTCGGTCAAACAAGGC
5' FGFR3-C	CGGCAGACGTACACGCTG
3' FGFR3-C	GTGGTGTGTTGGAGCTCATG
5' MMSET Exon 1	CCGAGGATGCGACGCACCGCAG
5' MMSET Exon 3	TGTCGAAGCAGCTCTTGTGT
CμB	GACGGAATTCTCACAGGAGAC
CδB	TGGGTGTCTGCACCCTGATAT
CγB	GGGGAAGACCGATGGGCCCT
CαB	GAGGCTCAGCGGGAAGACCTT
5' LETM1	GGAGTTTCTGCTGCCTGTTG
3' LETM1	ATCTCCTCGATGGTGTCCTG
5' RE-IIBP	GAGTAGCATTGTGGTTATAT
3' RE-IIBP	GATGCTGCCGTGGAAGCCCG

The sequence of primers used to detect beta-2-microglobin, der(4)t(4;14) IgH-MMSET hybrids, FGFR3, der(14)t(4;14) MMSET-IgH hybrids, LETM1, and RE-IIBP expression.

The single stage JH-ms6r and Iμ1-ms6r der(4) IgH-MMSET hybrid transcript screening reactions were carried out in 25 μl reactions using 50 ng of total RNA converted to cDNA as template. The reactions consisted of an initial 5 minute denaturation at 94^{O}C followed by 35 cycles of amplification at an annealing temperature of 60^{O}C and 1 minute extensions at 72^{O}C. The more sensitive two-stage der(4) IgH-MMSET hybrid transcript screening reactions utilized 1 μl of the Iμ1-ms6r product as template in a 50 μl reaction with the Iμ2 and ms5r primers followed by 30 cycles of amplification with identical cycling conditions as the single stage reactions. The der(14) MMSET-IgH hybrid transcript screening reaction conditions were the same as the one stage der(4) reactions except extension times were decreased to 30 seconds. *FGFR3* expression was detected by one-stage RT-PCR using the 5' and 3' FGFR3-C primers. The reaction conditions consisted of 30 cycles of amplification with an annealing temperature of 65^{O}C in a 50 μl reaction using 50 ng of total RNA converted to cDNA as template. In all cases, PCR products were resolved on 2% agarose gels and ethidium bromide stained DNA was visualized under a 254 nm UV lamp.

In general patients were screened for t(4;14) in batches of 22 along with a negative control and cDNA from the t(4;14) positive cell line, NCI-H929, diluted 1/10 as a positive control. Each set of screening reactions were only analyzed if the negative and positive controls were clean and positive, respectively. Subsequently each individual sample was scored as t(4;14) positive or negative based on the JH-ms6r, Iμ1-ms6r and the nested Iμ2-ms5r reactions if the beta-2-microglobulin cDNA integrity test was positive. The scoring criteria for *FGFR3* expression were identical. The diluted NCI-H929 cell line was chosen as the positive control specifically because it expresses the lowest amount of *FGFR3* among our t(4;14) positive/FGFR3 positive cell lines[88]. This ensures that patients with a minimal level of bone marrow plasmacytosis expressing low levels of *FGFR3* are detected as *FGFR3* expressors.

MM.3.4 - Quantitative RT-PCR Reactions

Gene expression levels were determined using qRT-PCR. Each reaction was performed in a volume of 50 µl with 1x TaqMan® Universal PCR Master Mix No AmpErase® UNG (Applied Biosystems), 2-2.5 µl of TaqMan® Assays-by-Design^{SM} or Assays-on-Demand™ (Applied Biosystems) primer and probe mixes, and 5 ng of RNA converted to cDNA as template. Prior to test reactions all primer and probe mixes were titrated with a paired glyceraldehyde-3-phosphate dehydrogenase (GAPDH) control to generate conditions validating the ΔΔCt analysis method described in ABI User Bulletin #2 (www.appliedbiosystems.com) with a minimum PCR efficiency of 92%. All reactions were performed on an ABI Prism® 7700 Sequence Detection System (Applied Biosystems). Primers used are as follows: Assays-on-Demand GAPDH (Hs99999905_m1), TACC3 (Hs00170751_m1), FGFR3 (Hs00179829_m1), LETM1 (Hs00360061_m1), MMSET Total (Hs00370212_m1), and WHSC2 (Hs00171805_m1); Assays-by-Design (Table M.2). Since the RE-IIBP assay could also amplify DNA we confirmed that no detectable DNA template was present with no-RT controls. No amplification was detectable in the absence of reverse transcribed cDNA.

Table MM.2 – Quantitative RT-PCR Primers and Probes

Oligo Name	Sequence (5'-3')
5' RE-IIBP	TGATATGTCTGGGATTTGCTTCAG
3' RE-IIBP	GCCCCCAAAAACATCAACCT
RE-IIBP Probe	CGTGGCTCTATCCATAC
5' Exon 4a	CACTCCTTCACAGCTATACCAAACTT
3' Exon 4a	GGGAGGAGAGAGGAAGGCA
Exon 4a Probe	TTGTTTTTAAGTGTTCAAACTTC
5' MMSET I	CATCTCCTTCTGCATCCTTAACTGA
3' MMSET I	AGCAGCTGGGTTCAAATCCAA
MMSET I Probe	CTCCCACAAAAGCTC
5' MMSET II	TCATTTGGGTGAAACTTGGGAACT
3' MMSET II	TCCAATCTCGTGCTTCATTTTCTGA
MMSET II Probe	CCGGCCACCATCTGT

Taqman® primer and probe pairs were designed by Applied Biosystems using the Assays-by-Design^{SM} service. All probes contain the 5' reporter dye 6-FAM and the 3' minor groove binder/non-fluorescence quencher (MGB/NFQ).

MM.3.5 – PCR Amplification of Genomic t(4;14) Breakpoints

Due to the large genomic region involved on both chromosomes, a comprehensive strategy was developed to predict the relative breakpoint regions on chromosomes 4 and 14. Each patient was assigned to specific breakpoint subgroups on chromosome 4 based on the products detected in the der(4) IgH-MMSET and der(14) MMSET-IgH hybrid transcript reactions to limit the number of PCR reactions. The specific clusters are defined as follows:

MB4-1a - Patients with predicted breakpoints telomeric of *MMSET* exon 1. These patients have MB4-1 type bands in the IgH-MMSET reactions, express *FGFR3*, and have no detectable MMSET-IgH hybrid transcripts.

MB4-1b - Patients with predicted breakpoints between *MMSET* exon 1 and *MMSET* exon 3. These patients have MB4-1 type bands in the IgH-MMSET reactions, express *FGFR3*, and have a detectable MMSET-IgH hybrid transcript.

MB4-1 - Patients with predicted breakpoints telomeric of *MMSET* exon 3. These patients have MB4-1 type bands in the IgH-MMSET reactions, but do not express *FGFR3*. As such the der(14) MMSET-IgH assays are deemed non-informative so no assumptions are made regarding the breakpoint location relative to *MMSET* exon 1.

MB4-2 - Patients with predicted breakpoints between *MMSET* exon 3 and *MMSET* exon 4. These patients have MB4-2 type bands in the IgH-MMSET reactions.

MB4-3a - Patients with predicted breakpoints between *MMSET* exon 4 and *MMSET* exon 4a. These patients have MB4-3 type bands in the IgH-MMSET reactions with a visible upper band corresponding to hybrid transcripts containing the alternatively spliced *MMSET* exon 4a.

MB4-3b - Patients with a predicted breakpoint between *MMSET* exon 4a and *MMSET* exon 5. These patients have MB4-3 type bands in the IgH-MMSET reactions with no visible doublet band corresponding to transcripts containing *MMSET* exon 4a.

In general we attempted to amplify the der(4) breakpoint first because this derivative is almost universally detected, whereas the der(14) appears to be lost in some patients[1,82]. The specific breakpoint defines the most centromeric *MMSET* exon primer used. The results of the Iμ1-ms6r and JH-ms6r reactions define which IgH primer will be used to amplify the breakpoint. In general both reactions are positive and thus the Iμ1 primer was used in most situations. However, some patients appear to have chromosome 14 breakpoints between the joining segment JH6 and Iμ as they are positive in JH-ms6r reactions but negative even in the nested Iμ2-ms5r reactions. In these rare cases, the der(4) breakpoints were amplified with sense primers residing in Eμ or JH6. The der(14) breakpoints were cloned based on the der(4) breakpoint sequence, the IgH isotype associated with der(14) MMSET-IgH hybrid transcripts (when detected), and/or the clinical IgH isotype.

Primers used to amplify the breakpoints are listed in Table M.3 and sub-grouped according to each breakpoint subtype. The breakpoints were amplified in 50 μl PCR reactions consisting of 60 mM Tris-SO_4, 18 mM Ammonium Sulfate, 2.0 mM $MgSO_4$, 200 μM of each dNTP, 2.5 U Platinum Taq DNA polymerase High Fidelity (Invitrogen), 0.2 μM of each PCR primer, and 150 ng of template DNA. The PCR reactions involved an initial denaturation at 96°C for 5 minutes followed by 35 cycles of amplification consisting of a 30 second denaturation at 94°C, a 30 second annealing at 60°C, and a 6 minute extension at 68°C followed by a final extension at 68°C for 20 minutes. The PCR products were resolved on 0.8% agarose gels and DNA bands were visualized under a 254 nm UV lamp with ethidium bromide staining. The PCR amplified bands were purified for sequencing using ExoSap-IT (Amersham Biosciences) or Genelute EtBr minus spin columns (Sigma). The purified products were directly sequenced with the Big Dye V1.1 cycle sequencing kit (Applied Biosystems) and the appropriate bi-directional tiling primers for each respective amplicon, listed in Table M.3. The sequencing reactions were run on an ABI 3100 genetic analyzer (Applied Biosystems) and the resulting sequence was analyzed with Sequence Analysis V3.7 (Applied Biosystems). The sequences were aligned to genomic contigs representing 4p16 and 14q32 using the pair-wise BLAST tool with the filter off to identify the involved regions and specific breakpoint sites.

Table MM.3 – PCR Primers used to Clone and Sequence t(4;14) Breakpoints

Subset	Primer	Short Name	Sequence (5’-3’)
Primers Used to Amplify der(4) Breakpoints			
5’ IgH	Switch8-Sense	Sw8-S	GCACCCACAGCAGGTG
	Imu1	Iμ1	AGCCCTTGTTAATGGACTTG
MB4-1a	3' LETM1 Intron 2	3’ LI2	CTTAGCCTCCCAAAGTGCTG
	3' LETM1 Intron 1.1	3’ LI1.1	ATTACAGGCGTGAGTTTGGG
	3' LETM1 Exon 1	3’ LE1	GGCGTCCATCTTACTGAGGA
	3' MMSET A	3’ MA	CAACTGTCCACTCCAGCTCA
	3' MMSET Exon 1a	3’ ME1a	AGTGTCCAGGCAGGAGAAGA
	3' MMSET Intron 1a	3’ MI1a	TAATCCCAGCACTTTGGGAC
	3' MMSET Exon 1	3’ ME1	CTGCGGTGCGTCGCATCCTCGG
MB4-1b	3' MMSET Intron 1.1	3’ MI1.1	TCTCAAGTGATCCTCCCACC
	3' MMSET Intron 1.2	3’ MI1.2	GCTCTCTGCTCCCTGCTCTA
	3' MMSET Intron 1.3	3’ MI1.3	GTTGGGAAGAAGAAGGGAGG
	3' MMSET Intron 1.4	3’ MI1.4	TAGAACCACTGGAACCCAGG
	3' MMSET Intron 1.5	3’ MI1.5	ATAGCAGACCAGGCACAGCT
	3' MMSET Exon 3	3’ ME3	AGCTTGTCGGCTGGAATAAA
MB4-2	3' MMSET Intron 3.1	3’ MI3.1	GAGCACTTCAGGAGGCAGAG
	3' MMSET Intron 3.2	3’ MI3.2	AGTCCTTGGGGCTCTCTGA
	3' MMSET Exon 4	3’ ME4	GGAGTGGATCTGCAGAAACC
MB4-3a	3' MMSET Intron 4.1	3’ MI4.1	TGTACCCCATTCCTAGCAGC
	3' MMSET Intron 4.2	3’ MI4.2	AGGCATGAGAGGCTAACAGC
	3' MMSET Intron 4.3	3’ MI4.3	AGAGATGGAGGCTAGGCACA
	3' MMSET Intron 4.4	3’ MI4.4	TGGTGAAACCCCAGCTCTAC
	3' MMSET Exon 4a	3’ ME4a	TCTGGGGAGGAGAGAGGAAG
MB4-3b	3' MMSET Intron 4a.1	3’ MI4a.1	AGATGTACAGGGCGGAACAT
	3' MMSET Intron 4a.2	3’ MI4a.2	CCTGGCTAACACGGTGAAAT
	3’ MMSET Exon 5	ms5r	AAGAACTGTACGTGATACTG

Primers Used to Amplify der(14) Breakpoints			
MB4-1a	5' LETM1 Exon 3	5’ LE3	CCGGAGCAGACTGTTAGAGC
	5' LETM1 Intron 2	5’ LI2	CAGCACTTTGGGAGGCTAAG
	5' LETM1 Intron 1.1	5’ LI1.1	CCCAAACTCACGCCTGTAAT
	5' LETM1 Intron 1.2	5’ LI1.2	GCCTTGGAGTTCAAGACCAG
	5' MMSET A	5’ MA	TGAGCTGGAGTGGACAGTTG
	5' MMSET Exon 1a	5’ ME1a	TCTTCTCCTGCCTGGACACT
MB4-1b	5' MMSET Intron 1a	5’ MI1a	GTCCCAAAGTGCTGGGATTA
	5' MMSET Exon 1	5’ ME1	CCGAGGATGCGACGCACCGCAG
	5' MMSET Intron 1.1	5’ MI1.1	GGTGGGAGGATCACTTGAGA
	5' MMSET Intron 1.2	5’ MI1.2	TAGAGCAGGGAGCAGAGAGC
	5' MMSET Intron 1.3	5’ MI1.3	CCTCCCTTCTTCTTCCCAAC
	5' MMSET Intron 1.4	5’ MI1.3	CCTGGGTTCCAGTGGTTCTA
	5' MMSET Intron 1.5	5’ MI1.5	AGCTGTGCCTGGTCTGCTAT
MB4-2	5' MMSET Exon 3	5’ ME3	TGTCGAAGCAGCTCTTGTGT
	5' MMSET Intron 3.1	5’ MI3.1	CTCTGCCTCCTGAAGTGCTC
	5' MMSET Intron 3.2	5’ MI3.2	TCAGAGAGCCCCAAGGACT
MB4-3a	5' MMSET Exon 4	5’ ME4	GGTTTCTGCAGATCCACTCC
	5' MMSET Intron 4.1	5’ MI4.1	GCTGCTAGGAATGGGGTACA
	5' MMSET Intron 4.2	5’ MI4.2	GCTGTTAGCCTCTCATGCCT
	5' MMSET Intron 4.3	5’ MI4.3	TGTGCCTAGCCTCCATCTCT
	5' MMSET Intron 4.4	5’ MI4.4	GTAGAGCTGGGGTTTCACCA
MB4-3b	5' MMSET Exon 4a	5’ ME4a	CTTCCTCTCTCCTCCCCAGA
	5' MMSET Intron 4a.1	5’ MI4a.1	ATGTTCCGCCCTGTACATCT
	5' MMSET Intron 4a.2	5’ MI4a.2	ATTTCACCGTGTTAGCCAGG
3’ IgH	CmuB	CμB	GACGGAATTCTCACAGGAGAC
	CdeltaB	CδB	TGGGTGTCTGCACCCTGATAT
	CgammaB	CγB	GGGGAAGACCGATGGGCCCT
	CalphaB	CαB	GAGGCTCAGCGGGAAGACCTT
Additional Primers used to Sequence Breakpoints			
5’ IgH	Imu2	Iμ2	CTTTGCAAGGCTCGCAGTGAC
	Switch14-Sense	Sw14-S	GGGTATCAAGTAGAGGGAGACA
3’ IgH	3' Switch-G1	Sw-G1	TCCCAGTGTCCTGCATTACTTC
	3' Switch-G2	Sw-G2	TTCTGGAGGCTCAGTCACCAC
	3' Switch-G3	Sw-G3	CTGGCAGCCCATCTTGC
	3' Switch-A1	Sw-A1	AACCCAGCACAGAGAGGCCT
	3' Switch-A2	Sw-A2	AGAGAGGCCTGGTGACAGCC

MM.4 – Analysis of MMSET Protein Variants

MM.4.1 - Expression Vector Construction

Plasmids containing the MMSET I and MMSET II ORF were kindly provided by Leif Bergsagel, Mayo Clinic (Scottsdale). With the intention of generating GFP fusion constructs in the Creator Cloning System (BD Biosciences, San Jose, CA), we amplified the required MMSET coding sequences from the provided plasmids by PCR with Platinum Taq DNA Polymerase High Fidelity (Invitrogen) and primers containing either NdeI or XhoI restriction sites (Table MM.4). The amplified products were sub-cloned in the pCR4.0-TOPO TA cloning plasmid (Invitrogen) and individual colonies were screened and sequenced to verify the restriction sites and flanking sequence.

Table MM.4 – Primers Used to Clone MMSET Variants

Primer	Sequence (5'-3')
5' MMSET (NdeI)	g**catatg**CTGGATGGAATTTAGC
5' MB4-2 (NdeI)	g**catatg**ATTCCAGCTAAGAAAGAG
5' MB4-3 (NdeI)	g**catatg**GTCAGAAAAAGAGTGCAC
5' RE-IIBP (NdeI)	g**catatg**TAAAATGATGCGGTG
3' MMSET I (XhoI)	g**ctcgag**CTAAGTGCAGTACAGAGC
3' MMSET II (XhoI)	g**ctcgag**CTATTTGCCCTCTGTC
3' MMSET Exon 4a (XhoI)	g**ctcgag**TTAATCTTTCAGTACAATTTG
3' MMSET I (Δstop-XhoI)	g**ctcgag**GTGCAGTACAGAGCAGC
3' MMSET II (Δstop-XhoI)	g**ctcgag**TATTTGCCCTCTGTGAC
3' Exon4a (Δstop-XhoI)	g**ctcgag**TCTTTCAGTACAATTTGAC
5' MMSET (XhoI)	gCGA**ctcgag**GCTGGATGGAATTTAG
3' Exon4a (Δstop-EcoRI)	gAGC**gaattc**AGTgcATCTTTCAGTACAAT

The various primers used to amplify the MMSET protein variants are listed. The translation initiation and termination codons are underlined. The restriction sites are highlighted in bold font. Lower case letters represent mismatched sequence required to introduce the restriction sites or to delete the stop codon. Any mismatch incorporated into the coding sequence did not change the encoded amino acid.

The NdeI/XhoI and NdeI/Δstop-XhoI cloned fragments were transferred by site directed cloning to pDNR-3 or pDNR-Dual, respectively, of the Creator™ Cloning System (BD Biosciences). The ORF of interest was then transferred from the donor vector into either pLP-EGFP-C1 or pLPS-3'EGFP using Cre Recombinase (BD Biosciences) following the manufacturer's instructions.

The alternatively spliced transcript, Exon 4a/MMSET III (NCBI Accession #AY694128), that includes exon 4a was cloned and sequenced from a t(4;14) positive myeloma patient with an MB4-1 breakpoint using the 5' MMSET (NdeI) and 3' MMSET Exon 4a (XhoI) primers and subsequently cloned into the Creator™ Cloning System as noted above. To create the in-frame fusion of Exon4a/MMSET III and B23 the Exon4a/MMSET III ORF was amplified with 5' MMSET (XhoI) and 3' Exon4a (Δstop-EcoRI) primers. Subsequently, the amplified ORF was cloned into the pEGFP-C1-B23 plasmid kindly provided by M. Hendzel, Cross Cancer Institute. Plasmids were prepared for transfection using the HiSpeed Plasmid Midi Kit (Qiagen, Mississauga, Canada).

MM.4.2 - Transfections, Immunoblotting and Microscopy

The HeLa cell line was transfected with each plasmid construct using Lipofectamine 2000 (Invitrogen). For immunoblot applications the transfections were performed in T25 flasks (Corning) while transfection for microscopy purposes were performed in 6 well plates with HeLa cells grown on coverslips. Protein expression was verified 24hrs post transfection by immunoblot. Cells were released from plates with 1x trypsin, washed 3x with PBS and lysed at $5x10^6$-10^7 cells/ml in 1% CHAPs plus 10 μg/mL leupeptin, 10 μg/mL antipain and 1 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich). 50 μl of lysate was run on a 5% stacking/8% separating SDS-PAGE gel. Fusion proteins were detected with a polyclonal anti-GFP serum (BD Biosciences).

All imaging of GFP tagged proteins was performed on live HeLa cells in conditioned culture media using an LSM 510 confocal microscope with a 40x/1.3 oil objective (Carl Zeiss, Thornwood, NY) and a Tempcontrol-mini (Zeiss) objective warmer set at 37^OC to minimize heat loss. To determine the *in vivo* localization, the pinhole was set to collect a 2 μm optical slice and 2 μg/ml of Hoechst 33342 (Molecular Probes, Eugene, OR) DNA stain was added to the media prior to imaging. Fluorescence Recovery After Photobleaching (FRAP) experiments were performed with the pinhole at max (1000 μm slice), a constant digital zoom of 4.5 (0.1 x 0.1 μm scaling), and a constant bleaching region of interest (ROI) height. The total imaging time and interval between image acquisitions was determined in pilot experiments for each construct. For each experiment the signal intensity for three different ROI's were collected using the LSM 510/Version 3.0 SP3 software package (Carl Zeiss); ROI1 the bleached region, ROI2 the entire nuclear fluorescence, and ROI3 the background fluorescence. The relative fluorescence intensity (RFI) for each timepoint was determined by first subtracting ROI3 from both ROI1 and ROI2 and then calculated as (ROI1timepoint/ROI1pre-bleach)/(ROI2timepoint/ROI2pre-bleach). The first image acquired after bleaching was set to an RFI of 0 and the corrected pre-bleach image was set to an RFI of 1.

Chapter 1

Frequency and Clinical Significance of t(4;14)(p16;q32) in Multiple Myeloma and Monoclonal Gammopathy of Undetermined Significance

C1.1.1 - Brief Introduction

Multiple Myeloma is a genetically unstable malignancy of post germinal center B lineage cells. The clonal cells of each patient are characterized by their unique clonotypic rearrangement of a variable, diversity, and joining (VDJ) gene segment on one immunoglobulin heavy chain (IgH) locus[15]. This clonotypic VDJ rearrangement represents a functionally rearranged IgH locus from which the monoclonal protein detected in the serum of most myeloma patients originates. The non-rearranged or non-functionally rearranged IgH locus on the alternative chromosome 14 is hypothesized to be the site of myeloma initiating translocations[231]. In myeloma these translocations typically occur in regions of the IgH loci that mediate class switch recombination, termed switch regions[73,89,232,233]. Since most myeloma patients produce a monoclonal protein, and the majority of the IgH translocations occur in regions downstream of the VDJ segments it is predicted that the chromosome 14 with the non-rearranged or non-functionally rearranged IgH locus is the one involved in the IgH translocations associated with myeloma[231].

Translocations involving the IgH loci are the most common genetic events in myeloma occurring in approximately 70-75% of patients[65,79]. Unlike some other hematological malignancies, no one specific translocation event characterizes all myeloma patients. A number of recurrent IgH translocations exist including; t(11;14)(q13;q32), t(4;14)(p16;q32), t(14;16)(q32;q23), t(6;14)(p25;q32), t(6;14)(p21;q32), and t(14;20)(q32;q12) which are predicted to dysregulate *CCND1*, *FGFR3/MMSET*, *MAF*, *IRF4*, *CCND3*, and *MAFB*, respectively[67]. However, the recurrent translocations are only found in approximately 40% of all myeloma patients[26].

The t(4;14) translocation is undetectable by conventional cytogenetics or spectral karyotyping due to the sub-telomeric localization of both chromosomal partner domains. It was first identified by Southern-blotting and subsequently by FISH and RT-PCR in MM, MGUS, and Primary Amyloidosis[83,88,91,96,101,108,119,234]. The genomic locations of t(4;14) breakpoints on chromosome 14 are almost exclusively in or near switch regions of the IgH loci[89,232]. The breakpoints on chromosome 4 occur within an approximate 65 kb region between *LETM1* exon 3 and *MMSET* exon 5 (See Chapter 4). This breakpoint region is a small part of a conserved gene cluster containing the *TACC3*, *FGFR3* and *MMSET* genes. As a result of the translocation *FGFR3* and *TACC3* may be dysregulated by the strong 3' IgH regulatory regions on der(14) while *MMSET* may be dysregulated by the intronic mu enhancer on der(4).

The initial t(4;14) frequency estimate of approximately 25% in multiple myeloma was generated from a large panel of myeloma cell lines and a very small cohort of patient samples[88]. When this project was initiated several small patient studies had identified a variable frequency from 12.6-20.8% [99,101-103]. Shortly after this project was completed, two contradictory reports regarding the survival of myeloma patients with t(4;14) were published with the larger study reporting a decrease in overall survival[83,119].

C1.2.1 – Specific Aim

To determine the incidence rate and clinical impact of t(4;14)(p16;q32) in multiple myeloma patients.

C1.3 – Chapter 1 Results

Almost all of the data presented in this chapter is published in Blood in one of two publications; Keats et al. 2003 or Keats et al. 2005[1,122]. The overall patient outcome data includes an additional 36 months of follow up compared to the initial data presented in Keats et al. 2003[1].

C1.3.1 – Sample Acquisition and Screening Assays

To determine the frequency of t(4;14) in multiple myeloma and MGUS we screened all available samples archived in our tumour bank. Our lab has maintained a tumour bank of BMMC, PBMC, and mobilized blood samples from patients with various gammopathies since October 1994. All patients with a diagnosed monoclonal gammopathy are invited to participate in our research studies so no selection bias exists in our acquisition of patient material. The only selection criterion for inclusion in the study was the availability of a bone marrow sample regardless of assumed diagnosis, disease course, and treatment. Therefore, all bone marrow samples archived in our tumour bank were screened for t(4;14). The diagnosis of each patient was independently determined from the available clinical data.

Several RT-PCR based assays have been used to screen patients for t(4;14). These include the der(4) IgH-MMSET and der(14) MMSET-IgH hybrid transcript assays, and the qualitative detection of *FGFR3* expression. We tested all of the assays with a series of 10 previously characterized t(4;14) positive myeloma cell lines to determine the most sensitive assay (Figure C.1). Both the der(4) IgH-MMSET hybrid transcript assays and the detection of *FGFR3* expression identified 90% of the t(4;14) positive cell lines. Importantly, the der(14) MMSET-IgH hybrid transcript assays only identified 40% of the t(4;14) positive cell lines. Based on these results, either the der(4) IgH-MMSET hybrid transcript assays or the detection of *FGFR3* expression appeared to be acceptable screening assays. We decided to use the der(4) IgH-MMSET hybrid transcript assays over the detection of *FGFR3* expression for several reasons. First, Malgeri et al. had recently compared the gold standard FISH assay with the JH-ms6r and Iμ1-ms6r der(4) IgH-MMSET hybrid transcript assays on 53 myeloma patients and showed a complete concordance[101]. Second, we intended to determine the clinical impact of t(4;14) and preferred a direct as opposed to an indirect measure of t(4;14) status. Third, the use of both the JH-ms6r and Iu1-ms6r assays ensures that all samples are independently tested for t(4;14) in at least two different replicates. Fourth, the der(4) hybrid transcript assays provide additional information on the approximate breakpoint location on chromosome 4, as they identify three different breakpoint regions, MB4-1, MB4-2, and MB4-3 based on the product size. The literature to date reports only one t(4;14) positive patient by FISH who was negative in the der(4) hybrid transcript assays[82]. Interestingly, a single patient was reported which is positive in the hybrid transcript assays but FISH negative[118]. Moreover, this patient expressed *FGFR3* suggesting it is a true t(4;14) positive.

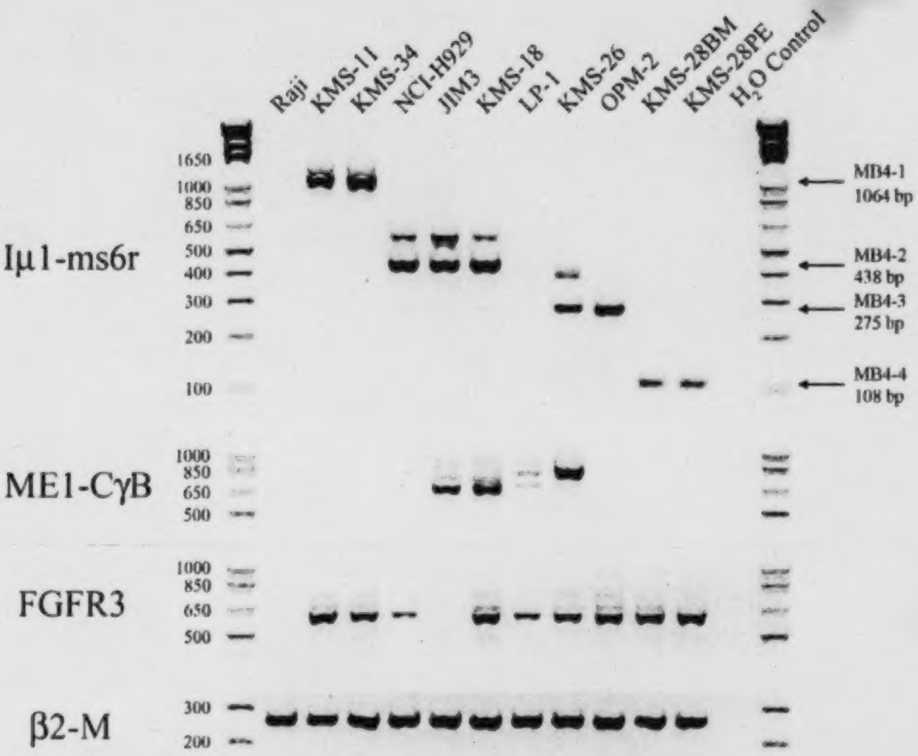

Raji
KMS-11
KMS-34
NCI-H929
JIM3
KMS-18
LP-1
KMS-26
OPM-2
KMS-28BM
KMS-28PE
H$_2$O Control
Iμ1-ms6r
1650
1000
850
650
500
400
300
200
100
MB4-1
1064 bp
MB4-2
438 bp
MB4-3
275 bp
MB4-4
108 bp
ME1-CγB
1000
850
650
500
FGFR3
1000
850
650
500
β2-M
300
200

Figure C1.1 – Test of Potential RT-PCR Based Screening Reactions

The various screening reactions were tested on a panel of available t(4;14) positive cell lines and the control cell line, Raji, to determine the most sensitive assay. The JH-ms6r (not shown) and the Iu1-ms6r reactions generated identical results. The der(14) ME1-CμB and ME1-CαB reactions are not shown, as all cell lines are negative in these reactions. The integrity of the tested cDNAs was verified with a beta-2-microglobulin specific RT-PCR. The four breakpoint clusters and their product sizes detected by the Iμ1-ms6r reaction are indicated to the right of the panel. The MB4-4 breakpoint seen in the KMS-28BM and KMS-28PE cell lines, results from an internal deletion on chromosome 4 that deletes *MMSET* exons 3-5.

The der(4) IgH-MMSET hybrid transcript assays were initially reported by Chesi et al. in 1998[91]. They were subsequent refined by Malgeri et al. in 2000 to increase the sensitivity of the assays[101]. The assays detect hybrid transcripts generated from der(4) which originate upstream of JH6 or Imu and subsequently aberrantly splice into the downstream exons of *MMSET* (Figure C1.2). Two features of the genomic breakpoints allow the assays to be highly specific and sensitive. First, the majority of breakpoints on chromosome 14 occur downstream of the Iμ region in switch mu or in downstream switch regions that have undergone class switch recombination with switch mu (Figure C1.2)[89]. Therefore, the large majority of breakpoints on chromosome 14 are within approximately 10 kb of Iμ either directly or as a result of class switch recombination. Second, all of the cloned t(4;14) breakpoints on chromosome 4 exist within a 65 kb region extending telomeric of the 5' end of the *MMSET* gene[89]. Therefore, since the telomeric IgH elements transferred to the der(4) are in the same transcriptional orientation as *MMSET* and the combined breakpoint region would in theory not be over 80 kb, the detection of hybrid IgH-MMSET transcripts with 5' primers in IgH segments with 3' primers in *MMSET* exons downstream of known breakpoints is quite robust. Furthermore, the assays define three major breakpoint regions, MB4-1, MB4-2, and MB4-3 based on the observed RT-PCR product sizes (Figure C1.2).

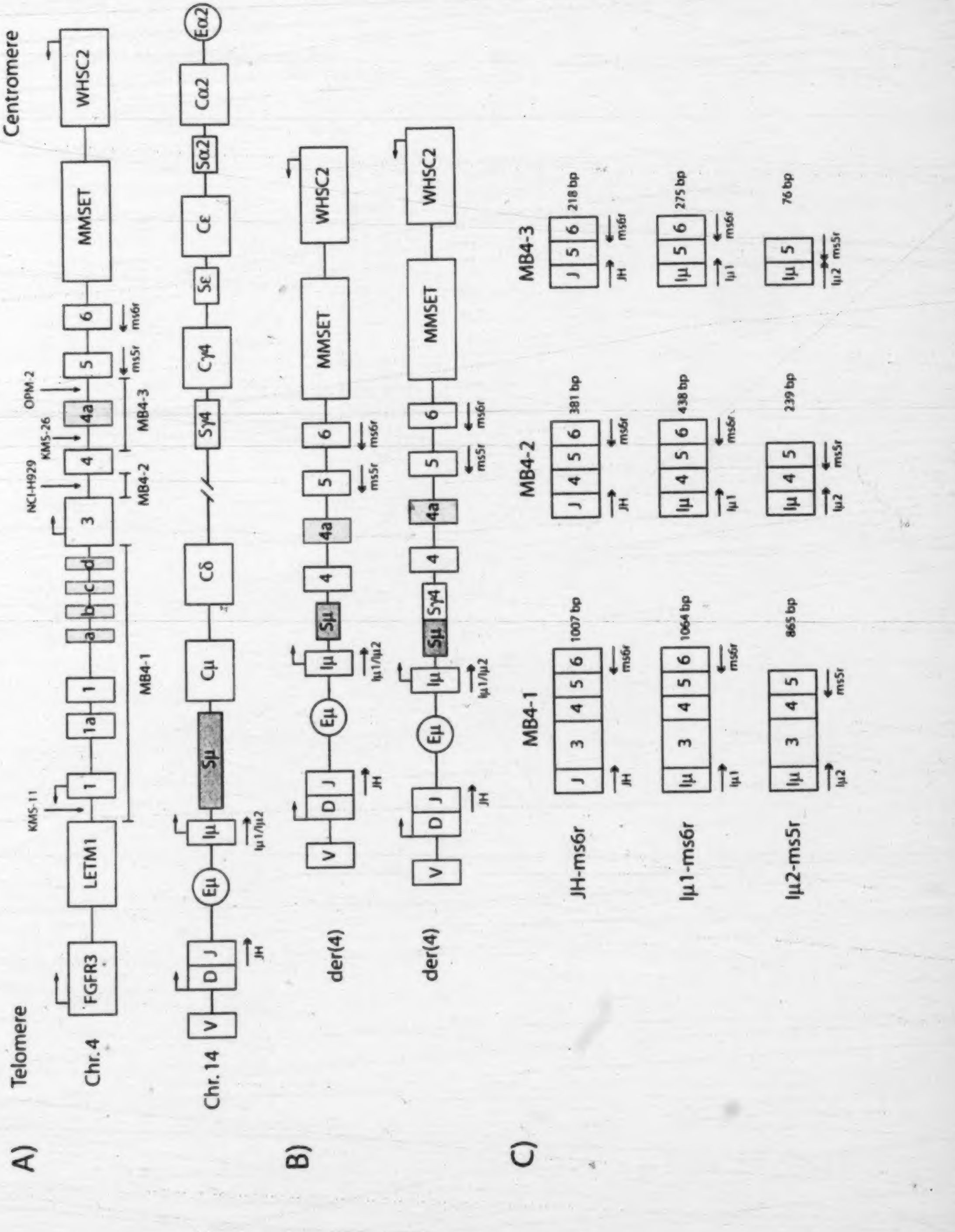

A)
Telomere
Centromere
Chr. 4
FGFR3
LETM1
KMS-11
1
1a
1
a
b
c
d
MB4-1
NCI-H929
3
MB4-2
KMS-26
4
OPM-2
4a
MB4-3
5
msSr
6
ms6r
MMSET
WHSC2
Chr. 14
V
D
J
JH
Eμ
Iμ
Iμ1/Iμ2
Sμ
Cμ
Cδ
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
B)
der(4)
V
D
J
JH
Eμ
Iμ
Iμ1/Iμ2
Sμ
4
4a
5
ms5r
6
ms6r
MMSET
WHSC2
der(4)
V
D
J
JH
Eμ
Iμ
Iμ1/Iμ2
Sμ
Sγ4
4
4a
5
ms5r
6
ms6r
MMSET
WHSC2
C)
MB4-1
MB4-2
MB4-3
JH-ms6r
J 3 4 5 6
JH
ms6r
1007 bp
J 4 5 6
JH
ms6r
381 bp
J 5 6
JH
ms6r
218 bp
Iμ1-ms6r
Iμ 3 4 5 6
Iμ1
ms6r
1064 bp
Iμ 4 5 6
Iμ1
ms6r
438 bp
Iμ 5 6
Iμ1
ms6r
275 bp
Iμ2-ms5r
Iμ 3 4 5
Iμ2
ms5r
865 bp
Iμ 4 5
Iμ2
ms5r
239 bp
Iμ 5
Iμ2
ms5r
76 bp

Figure C1.2 – Description of the der(4) IgH-MMSET Hybrid Transcript Assays

A) Diagram of the chromosomal regions involved in t(4;14) (not to scale). The individual *LETM1* and *MMSET* exons flanking the known breakpoints in t(4;14) positive cell lines are indicated and the alternatively spliced exons 2a, 2b, 2c, 2d, and 4a are highlighted in light grey. Vertical arrows indicate the relative breakpoint location on chromosome 4 of the cell lines characterizing each major breakpoint region. The proposed breakpoint regions on chromosome 4 for patients classified as MB4-1, MB4-2, or MB4-3 are indicated by boxed lines. The chromosome 14 diagram is a simplified representation of the IgH locus showing the predicted initial DJ rearrangement, a single telomeric VH segment to represent the VDJ region, and to further limit the size of the diagram the constant regions between IgD and IgG4 are not shown. The switch regions are indicated by small boxes with switch mu highlighted in dark grey and the enhancers are indicated by circles. The direction of transcription for each indicated gene is shown by directional arrows. The relative locations of the primers are indicated by lines with double arrowheads under the respective exon. B) Diagram of the two most common types of der(4) chromosomes created by t(4;14)(An MB4-2 breakpoint type is shown for simplicity). The upper diagram represents the most common rearrangement with switch mu joined directly to chromosome 4. The lower diagram represents the rare situations were a functional class switch recombination occurred with a subsequent translocation event involving the downstream switch region. C) Diagram of the various amplicons and product sizes detected in MB4-1, MB4-2, and MB4-3 patients with the der(4) IgH-MMSET hybrid transcript assays. For simplicity the alternatively spliced *MMSET* exon 4a is not shown.

The sensitivity of the der(4) IgH-MMSET hybrid transcript assays were determined by cell mixing experiments with the t(4;14) positive cell line, NCI-H929 (MB4-2), and the t(4;14) negative cell line U266. The one-stage assays using JH-ms6r or Iμ1-ms6r primer sets detect 1 in 50 positive cells (Figure C1.3). To increase the sensitivity further a nested two stage reaction was used in which the Iμ1-ms6r product was reamplified with the Iμ2-ms5r primer set. This assay was shown to readily detect 1 in 10 000 positive cells in the cell mixing experiments (Figure C1.3).

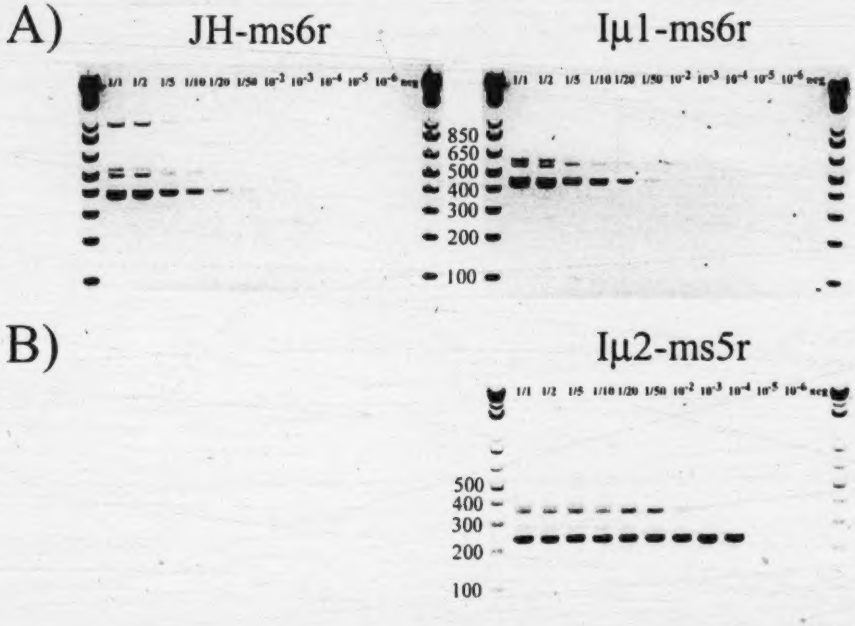

A)
JH-ms6r
Iμ1-ms6r
1/1 1/2 1/5 1/10 1/20 1/50 10-2 10-3 10-4 10-5 10-6 neg
1/1 1/2 1/5 1/10 1/20 1/50 10-2 10-3 10-4 10-5 10-6 neg
850
650
500
400
300
200
100
B)
Iμ2-ms5r
1/1 1/2 1/5 1/10 1/20 1/50 10-2 10-3 10-4 10-5 10-6 neg
500
400
300
200
100

Figure C1.3 – Sensitivity of the der(4) IgH-MMSET Hybrid Transcript Assays

A) Sensitivity of the single-stage JH-ms6r and Iμ1-ms6r assays is 1/50 positive cells as determined by a cell mixing experiment with the MB4-2 t(4;14) positive cell line, NCI-H929, and the t(4;14) negative cell line U266. B) Sensitivity of the nested Iμ2-ms5r two-stage assay is 1/10 000 positive cells.

C1.3.2 – The Occurrence of t(4;14)(p16;q32)

A total of 587 BM samples from 473 individuals collected between October 1994 and February 2004 were screened for t(4;14) with the der(4) IgH-MMSET hybrid transcript assays. This cohort includes 306 myeloma patients and 112 MGUS patients of which 44 (14.4%) and 2 (1.8%) are t(4;14) positive, respectively (Table C1.1). The t(4;14) positive myeloma patients are sub-grouped into 31 with the MB4-1 breakpoint, 6 with the MB4-2 breakpoint, and 7 with the MB4-3 breakpoint. A representative gel of each screening reaction indicating the various breakpoints is shown in Figure C1.4 and the screening results for each t(4;14) positive patient are listed in appendix I. Moreover, as the initial screening was independent of diagnosis, we screened; 9 patients with WM, 6 patients with primary amyloidoisis, 12 patients with plasmacytoma, and 28 individuals with no known malignancy (normal) or a malignancy unrelated to myeloma; and none were t(4;14) positive.

Table C1.1 – The Incidence of t(4;14)(p16;q32) in Myeloma and MGUS

Disease	Number of Patients Screened	Number of Patients with t(4;14)	Percent Positive
Myeloma	306	44	14.4%
		MB4-1 – 31 MB4-2 – 6 MB4-3 – 7	70.5% 13.6% 15.9%
MGUS	112	2	1.8%
		MB4-1 – 2	100%

Patients were screened for t(4;14)(p16;q32) with the der(4) IgH-MMSET hybrid transcript assays

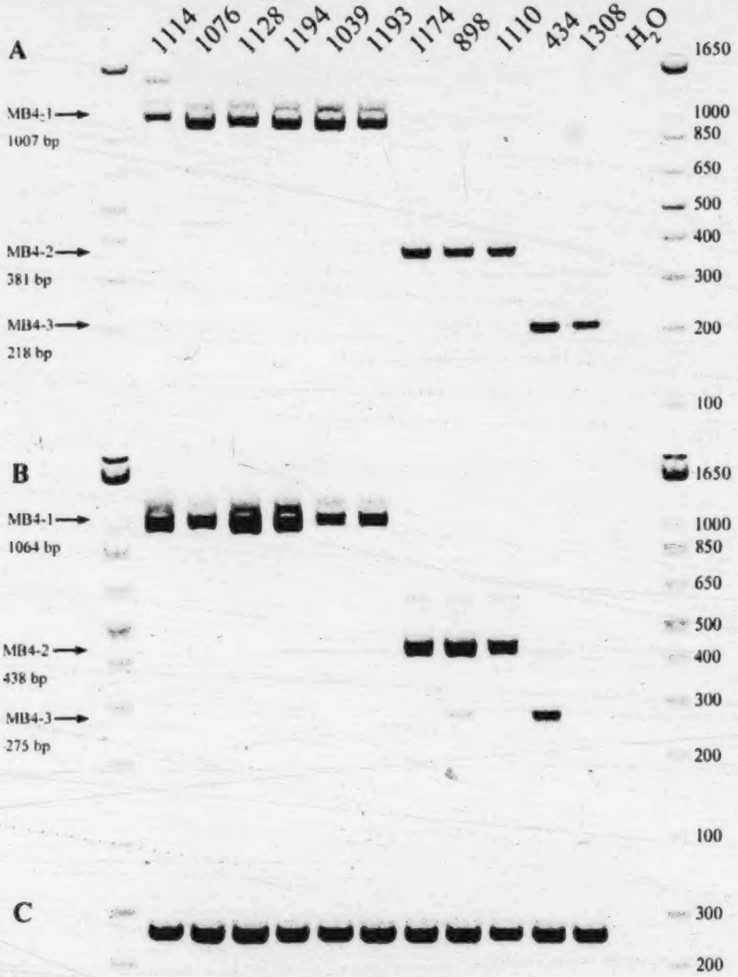

A
1114
1076
1128
1194
1039
1193
1174
898
1110
434
1308
H2O
1650
MB4-1 →
1007 bp
1000
850
650
500
400
MB4-2 →
381 bp
300
MB4-3 →
218 bp
200
100
B
1650
MB4-1 →
1064 bp
1000
850
650
500
MB4-2 →
438 bp
400
300
MB4-3 →
275 bp
200
100
C
300
200

Figure C1.4 – Representative Panel of the t(4;14) Positive Patients

A) Representative gel of eleven t(4;14) positive patients, with various breakpoints, detected by a JH-ms6r assay. B) Representative gel of the same eleven patients detected by the Iμ1-ms6r assay. C) cDNA integrity tests with beta-2-microglobulin primers on the eleven patient samples. Patient 1308 is one of the rare patients with an apparent genomic breakpoint on chromosome 14 between JH6 and Imu as this patient only has detectable der(4) IgH-MMSET transcripts with JH-ms6r and not with Iμ1-ms6r or the nested Iμ2-ms5r reaction.

To determine if t(4;14) could be a progression event in myeloma we screened sequential BM samples from patients at different disease stages. The majority of the samples were procured at diagnosis or relapse; however, several were scheduled as a part of various clinical trials. Sequential BM samples were available from 80 of the myeloma patients with a median number of 2 samples each (range, 2-5). The 80 patients included 66 t(4;14) negative and 14 t(4;14) positive .patients at diagnosis. The median time interval between the first and last sequential sample was 14.5 months (range, 1-84 months). All of the t(4;14) positive patients remained positive for hybrid transcripts over their respective timelines while the t(4;14) negative patients remained negative. Therefore, t(4;14) does not appear to be a progression event as no patient gained the translocation during their observed clinical course.

Of the 44 t(4;14) positive myeloma patients, 2 were previously diagnosed with MGUS. We have not been able to determine if these patients gained t(4;14) during their transition from MGUS to myeloma or if they were t(4;14) positive when diagnosed with MGUS; BM samples from these time points were not available. Assuming both patients were originally t(4;14) positive the transition is not immediate. Patient 1091 transitioned from MGUS to overt myeloma after 55 months while patient 830 transitioned from MGUS to asymptomatic myeloma after 74 months and continues to be untreated 40 months later. Furthermore, neither of the t(4;14) positive patients diagnosed with MGUS developed overt myeloma to our knowledge. However, patient 1653 has not been followed up clinically in the 17 months since the diagnosis of MGUS and patient 1307 passed away 30 months after the diagnosis of MGUS from what are believed to be unrelated issues. Finally, 5 of the 44 t(4;14) positive myeloma patients can be defined as having smoldering multiple myeloma or asymptomatic myeloma.

C1.3.3 – The Clinical Impact of t(4;14)(p16;q32)

Our initial cohort of patients consisted of BM samples from 208 myeloma patients collected between October 1994 and December 2001[1]. As shown in Table C1.2 this cohort of patients consists of 31 (14.9%) t(4;14) positive patients. The 31 t(4;14) positive patients subdivide into 19 with the MB4-1 breakpoint and 5 patients each with the MB4-2 or MB4-3 breakpoints. The baseline characteristics of the 208 myeloma patients are listed in Table C1.2. A total of 14 patients in the cohort have not received systemic therapy, either because they were diagnosed in terminal stages (n=7, including 2 t(4;14) positive cases) or because they have not required it (n=7, including 2 t(4;14) positive cases). Sixty-three patients were treated initially with cyclical vincristine, adriamycin, dexamethasone (VAD) chemotherapy followed by high dose melphalan and autologous stem cell transplant. One patient received an allogeneic transplant. Of the 126 patients listed in Table C1.2 as being treated with chemotherapy, 116 patients were treated principally with conventional chemotherapy regimens (85 with melphalan plus corticosteroids, 31 with VAD). Three of these patients were receiving VAD in preparation for autologous transplant. The remaining patients were treated with a variety of therapies (dexamethasone alone, n=6; delayed autologous transplant, n=2; single agent melphalan, n=1; cyclophosphamide plus dexamethasone, n=1; unspecified, n=4).

Table C1.2 – Clinical Characteristics of Patients in the Original Cohort

		Overall	t(4;14) Negative	t(4;14) Positive
N (%)		208	177 (85%)	31 (15%)
		median (range)	**median (range)**	**median (range)**
Age (yrs)		67 (30-89)	67 (30-89)	65 (41-87)
Sex		67 female	57 female	10 female
		141 male	120 male	21 male
		No. (%)	**No. (%)**	**No. (%)**
Durie-Salmon Stage	I	24 (12)	18 (10)	6 (19)
	II	32 (15)	26 (15)	6 (19)
	III	152 (73)	133 (75)	19 (61)
	B (sCr>2 mg/dL)	38/202 (19)	34/171 (20)	4/31 (13)
Clinical Isotype	IgG	119 (57)	100 (56)	19 (61)
	IgA	44 (21)	35 (20)	9 (29)
	light chain	34 (16)	32 (18)	2 (6)
	Other/unspecified	11 (5)	10 (6)	1 (3)
Lytic bone disease		136/205 (66)	117/176 (66)	19/29 (66)
Calcium > 12 mg/dL		18/198 (9)	16/168 (10)	2/30 (6)
Hgb < 8.5 g/dL		37/206 (18)	32/175 (18)	5/31 (16)
Elevated LDH		24/152 (16)	21/127 (17)	3/25 (12)
Beta-2-Microglobin > 3 mg/L		111/155 (72)	96/134 (72)	15/21 (71)
Urine protein > 4g/day		31/119 (26)	24/104 (23)	7/15 (47)
Primary Systemic Therapy	None	14 (7)	10 (6)	4 (13)
	Chemo	126 (61)	108 (61)	18 (58)
	Auto SCT	63 (30)	55 (31)	8 (26)
	Allo SCT	1 (0.5)	1 (0.6)	0
	Unspecified	4 (2)	3 (2)	1 (3)
Patients deceased		152 (73)	128 (72)	24 (77)
Patients Lost to follow-up		7 (3)	5 (3)	2 (6)
		median (range)	**median (range)**	**median (range)**
Total follow-up, years		3.1 (0-13.3)	3.3 (0-13.3)	1.8 (0-9.0)
Survivor follow-up, years		4.4 (0-11.1)	4.5 (0-11.1)	3.8 (0-6.9)

SCT indicates stem cell transplantation; sCR, serum creatinine; and Hb, hemoglobin. Patients are defined as being lost to follow-up if no record of the patients' vital status exists for over two years.

In this cohort, the presence of t(4;14) was not associated with statistically significant differences in any of the baseline patient factors examined (age, sex, clinical isotype, Durie-Salmon stage, beta-2-microglobulin, LDH, albumin, serum creatinine, calcium, hemoglobin, serum or urine monoclonal protein levels, % plasma cells in the bone marrow, the presence of circulating plasma cells on peripheral blood smears, the presence of lytic bone disease, and the use of high-dose chemotherapy). A trend towards increased proteinuria in t(4;14) positive patients was suggested by the data (Table C1.2), but this did not reach statistical significance (p=0.06) and was based on small patient numbers.

One hundred and thirty-five patients had sufficient data to be analyzed retrospectively for response to first-line therapy. A response was defined according to the National Cancer Institute of Canada Clinical Trials Group criteria[235]. The main reasons for exclusion included too few courses of therapy, or insufficient data found on the retrospective chart review. Of the 135 patients, 11/20 (55%) t(4;14) positive cases responded, compared to 89/115 (78%) t(4;14) negative cases (p=0.05). This trend towards reduced response to therapy in t(4;14) positive patients was seen in all treatment subgroups, but the subgroups were too small to reach statistical significance.

The median survival from the date of diagnosis in the 31 myeloma patients with t(4;14), as detected by the der(4) IgH-MMSET hybrid transcript assays, was 709 days. The median survival for the 177 patients with a negative RT-PCR result is 1338 days (Figure C1.5). The difference in survival between the two groups is statistically significant (P=0.037; HR=1.58, 95% CI 1.04-2.93).

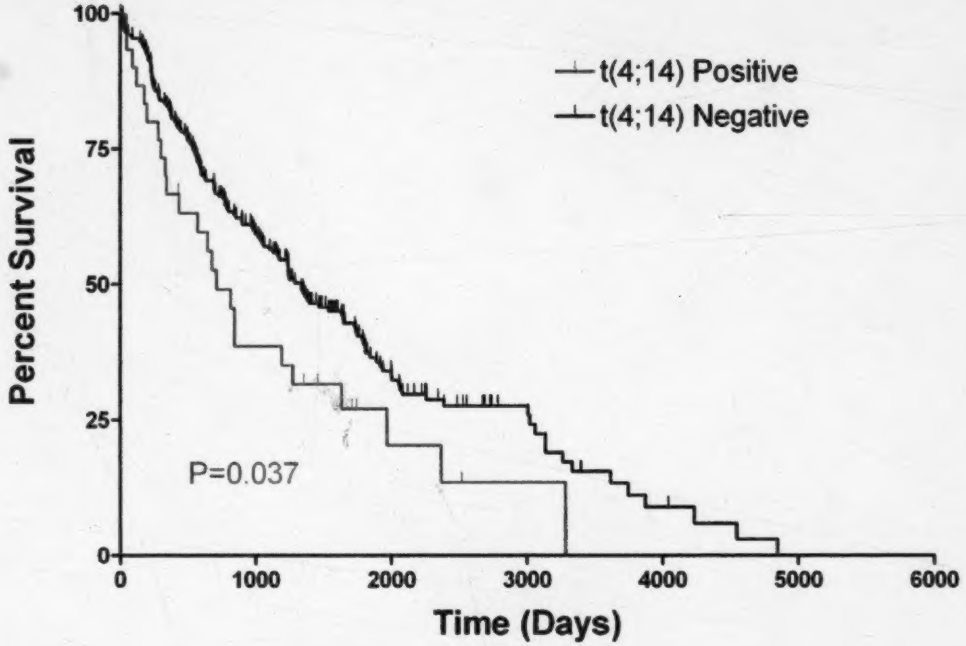


Figure C1.5 – Kaplan-Meier Overall Survival Plot for the Original Patient Cohort

This overall survival plot includes an additional 36 months of follow-up compared to the original analysis presented in Keats et al. 2003[1]. The black line indicates data for 177 t(4;14) negative patients with a median survival of 1338 days; red line indicates data for 31 t(4;14) positive patients with a median survival of 709 days. The difference in survival between the two groups was determined with the log rank test (HR=1.58; 95% CI 1.04-2.93; P=0.037). The cohort contains 7 patients lost to follow-up (no updated survival information for more than 2 years) of which 5 are t(4;14) negative and 2 are t(4;14) positive.

We used a Cox model to adjust the relationship between t(4;14) and survival for baseline patient characteristics and for the type of therapy received (i.e. standard dose versus high dose chemotherapy). Significant predictors of survival in the model included t(4;14) status (P=0.002), age (P=0.02), beta-2-microglobulin level (P<0.0001), BM plasmacytosis (P=0.01) and high-dose chemotherapy (P<0.001). The adjusted hazard ratio in t(4;14) positive patients was 2.6 (95% CI, 1.4-4.8).

This study was population-based, and was comprised of a diverse array of patients treated with a variety of therapies. In addition, not all patients provided samples at the time of diagnosis and it was assumed based on our longitudinal analysis (C1.3.2) that t(4;14) status does not change with time. However, the adverse prognostic effect of t(4;14) was significant in the subset of 137 patients who provided a diagnostic sample (P=0.01), even after adjusting for other prognostic factors such as high-dose chemotherapy in a Cox model. Furthermore, the adverse prognostic effect is maintained in the subgroup of 112 patients who received conventional chemotherapy (P=0.01).

C1.3.4 – Monitoring Disease Burden with IgH-MMSET Hybrid Transcript Assays

To determine if cells harboring the t(4;14) translocation are detectable outside of the BM we screened a random subset of patients from the original cohort with peripheral blood (PB) samples taken at the same time as the screened BM samples. Cells with the translocation were detected using the one-stage der(4) IgH-MMSET hybrid transcript screening assays. Of the 35 patients screened, 10 were identified as t(4;14) positive in the BM screening. Eight patients had t(4;14) detectable in the PB, all of whom were positive in the BM with the same breakpoint type. For 2 of the 10 patients, the PBMC samples screened were negative even though the BM samples were positive. Thus 80% of the t(4;14) positive patients had detectable cells in the PB. For the majority of these patients, t(4;14) positive cells in the total PBMC must be relatively frequent (>2% of PBMC) as the sensitivity of the single stage assay is 1 in 50 cells (Figure C1.3).

Since, der(4) IgH-MMSET hybrid transcripts are detectable in the peripheral blood of most t(4;14) positive patients at diagnosis, we determined the relationship between the detection of IgH-MMSET transcripts and disease burden, as measured by M protein levels and the detection of the clonotypic VDJ rearrangement, over the course of disease. We performed longitudinal analysis on 5 of the t(4;14) positive patients for whom we had sequential BMMC and PBMC samples covering the entire disease course and patient specific complement determining region 2 and 3 (CDR2/CDR3) primers. The der(4) IgH-MMSET transcripts were detectable in the peripheral blood at various points during the course of disease for each patient analyzed. Most often hybrid transcripts were detectable at diagnosis, became undetectable during therapy related remission and subsequently became detectable at relapse and were maintained during the terminal phase (Figure C1.6). The expression of *FGFR3* was also followed through the time courses. *FGFR3* expression often correlated with the detection of hybrid transcripts; however, in some situations only *FGFR3* or IgH-MMSET hybrid transcripts were detected at specific timepoints (Figure C1.6). Finally, we attempted to determine if the detection of IgH-MMSET hybrid transcripts was equivalent to the detection of clonotypic CDR2/CDR3 transcripts in monitoring minimal residual disease. In all patients analyzed, monitoring for clonotypic CDR2/CDR3 transcripts was a considerably more sensitive means to detect residual or emerging disease compared to monitoring for IgH-MMSET or FGFR3 transcripts (Figure C1.6).

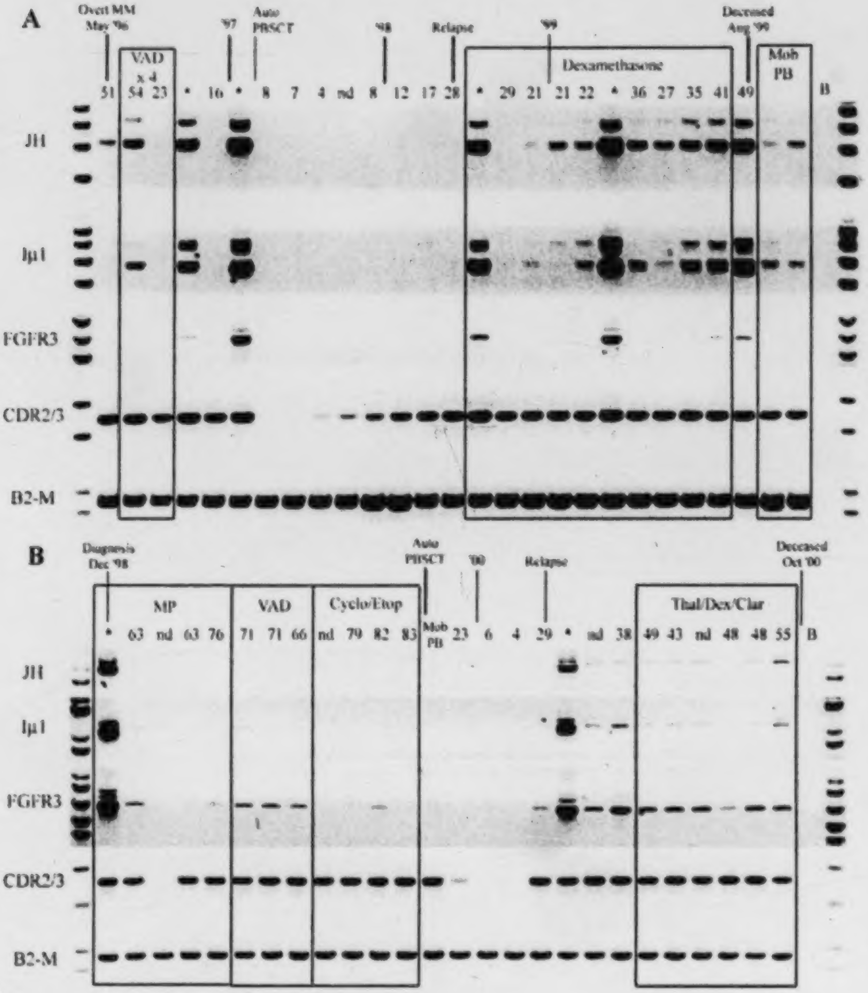

A
Overt MM
May '96
VAD
x 4
'97
Auto
PBSCT
'98
Relapse
'99
Dexamethasone
Deceased
Aug '99
Mob
PB
51 54 23 * 16 * 8 7 4 nd 8 12 17 28 * 29 21 21 22 * 36 27 35 41 49
B
JH
Iμ1
FGFR3
CDR2/3
B2-M
B
Diagnosis
Dec '98
MP
VAD
Cyclo/Etop
Auto
PBSCT
'00
Relapse
Thal/Dex/Clar
Deceased
Oct '00
* 63 nd 63 76 71 71 66 nd 79 82 83
Mob
PB
23 6 4 29 * nd 38 49 43 nd 48 48 55 B
JH
Iμ1
FGFR3
CDR2/3
B2-M

Figure C1.6 – Timeline Analysis of IgH-MMSET Hybrid Transcripts

PBMC and BMMC samples collected over the course of disease from two t(4;14) positive *FGFR3* expressing myeloma patients are shown. The monoclonal protein values (g/L) at each visit are marked above each PBMC sample and BMMC samples are denoted by an asterisk (*). Clinical diagnosis and relapse along with therapy are noted (Auto PBSCT, autologous peripheral blood stem cell transplant following high dose therapy; Mob PB, apheresis product from mobilized peripheral blood). Representative panels for each reaction are shown; JH, JH-ms6r hybrid transcript assay; Iμ1, Iμ1-ms6r hybrid transcript assay; FGFR3, *FGFR3* expression; CDR2/3, CDR2-CDR3 patient specific transcript assay; beta-2-microglobulin (B2-M), positive control for integrity of cDNA. (A) The patient shown, Edm 434, has the MB4-3 breakpoint type, however, the breakpoint occurs between exon 4 and exon 4a creating a doublet band. The upper band of the doublet represents alternative splicing to exon 4a. (B) The patient shown, Edm 1128, has the MB4-1 breakpoint type and expresses an *FGFR3* allele with the TDI associated mutation, Arg248Cys.

C1.3.5 – Unique Observations from the der(4) IgH-MMSET Screening Assays

First, we identified 3 patients (1207, 1308, and 1607) with detectable JH-MMSET hybrids but not Iμ-MMSET hybrids. A representative set of reactions from patient 1308 is shown in Figure C1.4. In all three patients this phenomena was detected at diagnosis and thus can not be attributed to a therapy induced downregulation of transcription from Imu. Moreover, this does not represent a sensitivity issue as the nested Iμ2-ms5r reaction should be positive given the level of bone marrow plasmacytosis in the three patients was 50%, 20%, and 23%. No obvious clinical correlations exist between these patients, however, the poor prognosis associated with t(4;14) appears to hold as patients 1207 and 1607 only survived 43 and 399 days, respectively. The remaining patient is still alive 1354 days post-diagnosis. This apparently unique breakpoint event is not associated with a specific sub-group as patients 1207 and 1308 have MB4-3 breakpoints and 1607 has an MB4-1 breakpoint. The only feature common to all three patients is the lack of detectable *FGFR3* expression (See Chapter 2). Based on the available data we hypothesize that the genomic breakpoints in these patients are in the intervening region between JH6 and the 3' splice site within Imu or alternatively in a downstream switch region. We have confirmed the former hypothesis in patient 1308 as the cloned der(4) breakpoint is between the mu enhancer and Imu (See Chapter 4).

Second, in patient 1223 the JH-ms6r and Iμ1-ms6r reactions produced a novel banding pattern with strong bands below the proper banding size for patients with MB4-3 breakpoints and weak bands at the proper size for MB4-2 bands (Figure C1.7). Unique bands originating from the der(4) IgH-MMSET hybrid transcript assays have been reported by other groups[119,236]. In one case, the abnormal banding pattern was proposed to be the result of an alternative splicing event between the middle of *MMSET* exon 3 and the middle of *MMSET* exon 5[119]. Although the cryptic splice sites were identified in an identical 13 bp repeat present in both exons, the presence of the intervening region was not verified. Therefore, it is not known whether this represents a genomic deletion due to a recombination between the 13 bp repeat regions or an alternative splicing event. In the other case, the abnormal band is detected in the KMS-28BM and KMS-28PE HMCLs (Figure C1.1)[236]. This band size is labeled as MB4-4 as it represents a hybrid transcript composed of JH or Imu spliced directly to *MMSET* exon 6 exclusively. For these cell lines the genomic breakpoint is in the MB4-1 region but a downstream internal deletion between *MMSET* intron 1 and *MMSET* intron 5 results in the novel MB4-4 band. In the case of patient 1223, the strong lower bands are the same size as the predicted MB4-4 bands, but the co-existence of weak bands corresponding to an MB4-2 breakpoint makes this a unique patient. To determine if the MB4-2 bands represent true RT-PCR products, we attempted to amplify the hybrid transcripts with JH-ms5r or Im1-ms5r. This produced clear bands of the predicted size for an MB4-2 positive patient (Figure C1.7). The strong bands identified in the Iμ1-ms6r and Iμ1-ms5r reactions were sub-cloned and sequenced. The sequencing results confirmed the predicted joining of IgH elements to *MMSET* exon 6 in the predicted MB4-4 band and the joining of IgH elements to *MMSET* exon 4 in the MB4-2 band. Several potential mechanisms can be proposed to explain the genetic events underlying this phenomenon. First, the translocation may have occurred twice with different breakpoints occurring each time. If this is the case, the translocation either reoccurred on the same IgH locus, both events occurred after a duplication of the non-functionally rearranged IgH locus, or the second event occurred as a secondary event on the functionally rearranged IgH locus of a subclone, since this patient produced a monoclonal protein. Secondly, this could represent a clonally heterogeneous patient with an initial MB4-2 genomic breakpoint and a subsequent deletion of MMSET exons 4 and 5 in a subset of the malignant clone. Third, this may represent a complex alternative splicing event; it is however difficult to explain why both MMSET exons 4 and 5 would be skipped in this situation. We attempted to verify the previous model by cloning the predicted MB4-2 breakpoint; however, to date we have not been successful (See Chapter 4).

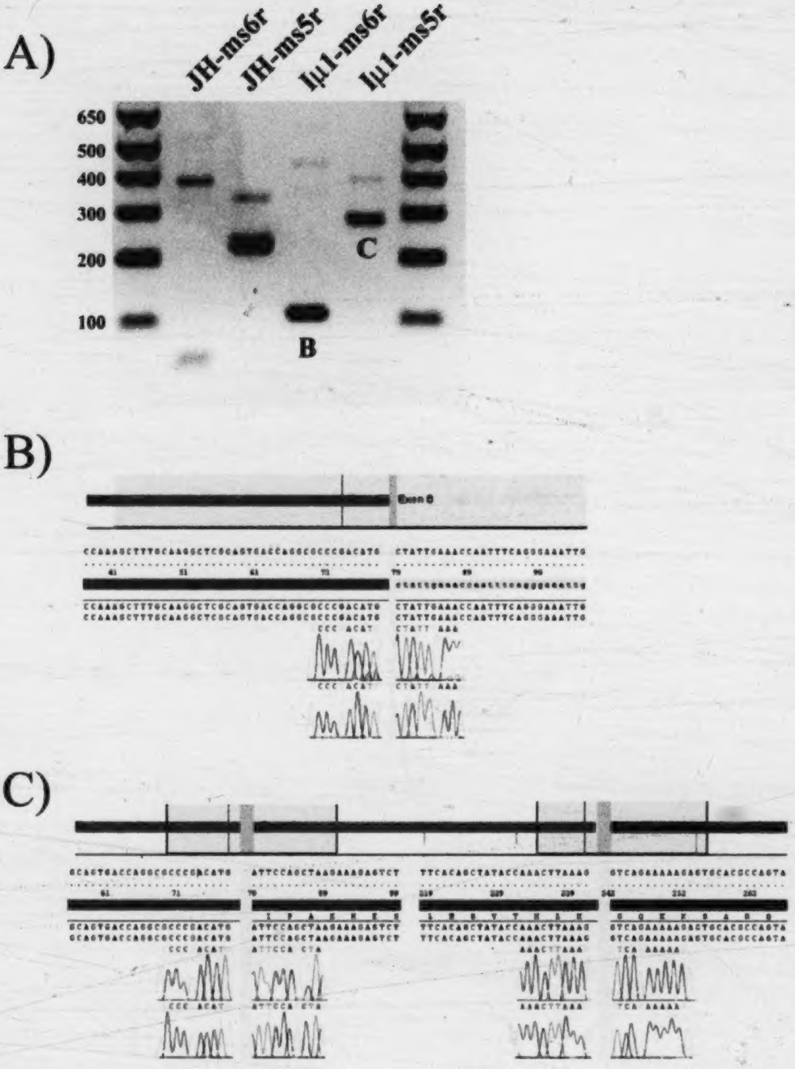

A)
JH-ms6r
JH-ms5r
Iµ1-ms6r
Iµ1-ms5r
650
500
400
300
200
100
C
B
B)
Exon 8
C)

Figure C1.7 – Patient with both MB4-2 and Novel MB4-4 Hybrid Transcripts

A) Agarose gel with the RT-PCR products from JH-ms6r, JH-ms5r, Iµ1-ms6r, and Iµ1-ms5r reactions. The bands above letters B and C were subcloned and the resulting sequence is shown in panels B and C, respectively. B) Sequence of the lower Iµ1-ms6r reaction band. C) Sequence of the lower band from the Iµ1-ms5r reaction.

C1.4.1 – Chapter Conclusions

The principle goals of this study were to determine the frequency and associated clinical outcome of t(4;14) in a large cohort of myeloma patients. To a lesser extent we wished to determine if we could monitor the level of disease burden by monitoring the translocation with der(4) IgH-MMSET hybrid transcript assays or the detection of *FGFR3* expression.

The initial frequency estimate of approximately 25% was based on the frequency observed in a panel of human myeloma cell lines and a small patient subset[88]. In our cohort of patients with multiple myeloma 44/306 (14.4%) have the translocation as detected by the der(4) IgH-MMSET hybrid transcript assays. Therefore, the frequency of t(4;14) in a large patient population is approximately 10% lower than the frequency observed in the cell lines. Interestingly, the frequency of t(4;14) in MGUS is much lower, 2/112 (1.8%), suggesting the translocation promotes a myeloma phenotype.

In our original cohort of 208 myeloma patients with 31 t(4;14) positive patients the translocation predicts a poor prognosis with a median survival of 709 days compared to 1388 days in t(4;14) negative cases. Furthermore, the poor prognosis associated with t(4;14) is independent of known prognostic markers. Therefore, t(4;14) is a novel independent prognostic marker which predicts for a poor overall outcome.

The use of RT-PCR to detect JH-ms6r, Iμ1-ms6r, or even *FGFR3* transcripts to monitor tumour burden was not as sensitive as the patient specific CDR2-CDR3 assays. The patient specific assays identified relapsing disease much sooner than the hybrid transcript assays and more accurately reflected the clinical monitoring of the serum monoclonal protein. Therefore, to accurately monitor minimal residual disease or relapsing disease the use of patient specific CDR2-CDR3 assays are recommended. However, to quantitatively measure disease burden assays against the clonal VDJ rearrangement or the genomic t(4;14) breakpoint should provide equal sensitivity. Furthermore, it may prove easier to design quantitative assays against the genomic t(4;14) breakpoint compared to the clonal VDJ rearrangement due to similarity between variable, diversity and joining IgH segments.

Chapter 2

Target Genes and the der(14) Chromosome

C2.1.1 - Brief Introduction

Many hematological malignancies are characterized by unique, and often diagnostic, translocations. Multiple myeloma has no unique translocation although translocations involving the immunoglobulin heavy chain locus on chromosome 14 and multiple partner chromosomes are present in 70-80% of patients[65,66,83]. The translocation mechanism appears to be linked to the class switch recombination process, as most breakpoints are located in IgH switch regions[89,232,233]. The recurrent translocations t(11;14)(q13;q32), t(4;14)(p16;q32), and t(14;16)(q32;q23) are cumulatively present in approximately 40% of patients[67,79]. These translocations also predict a differential overall outcome with t(11;14) positive patients having an improved prognosis while t(4;14) positive and t(14;16) positive patients have a worse prognosis compared to patients without these genetic events[1,26,80,83].

The IgH switch translocations in myeloma separate the strong 3' IgH regulatory regions and the mu enhancer of the IgH locus onto different derivative chromosomes. For t(4;14), this results in the expression of *FGFR3* by the 3' IgH regulatory regions, while the mu enhancer is thought to increase the expression of *MMSET* [91]. This coordinate dysregulation of at least two genes makes the identification of a true target gene difficult. However, in approximately 30% of our patients, the breakpoints are downstream of the proper MMSET translation initiation site, making the potential contribution of *MMSET* unclear.

The 4p16 genomic region involved in t(4;14) is linked with a number of human genetic disorders including Huntington Disease, Wolf-Hirschhorn Syndrome (WHS), and the autosomal dominant skeletal disorders (hypochondroplasia, achondroplasia, thanatophoric dysplasia type I and II). The skeletal disorders are linked to activating mutations of *FGFR3*[95], and some of these mutations are found in 5-10% of t(4;14) positive patients[118,169,171]. The minimally deleted WHS critical region of 165 kb contains two genes; *WHSC1/MMSET/NSD2* and *WHSC2*[90,187,223]. Two different transcripts are transcribed from the *MMSET* gene. The first originates upstream of the proper translation initiation site and is alternatively spliced into three mRNA species which encode the MMSET type I, MMSET type II, and MMSET type III protein variants[90,91,101,122]. MMSET II, the full length protein, contains a SET domain and thus is predicted to regulate gene expression by modifying histone methylation patterns[237]. The second transcript originates within intron 9 of the *MMSET* gene[188]. Translation of this transcript initiates in exon 15 and produces the RE-IIBP protein, which is identical to the C-terminus of MMSET II.

Though *FGFR3* has been studied extensively as the t(4;14) target gene[35,129,130], the limited number of activating mutations in patient samples makes it questionable if this gene is the sole t(4;14) target gene. Furthermore, with all of the proposed alternative t(4;14) target genes on both derivative chromosomes it is increasingly likely that other genes are directly dysregulated by t(4;14). Therefore, it is essential that a comprehensive study determine the dysregulation of each gene in *ex vivo* patient samples so that the factors contributing to t(4;14) mediated myelomagenesis are identified.

C2.2.1 - Working Hypothesis

1) *That FGFR3 and MMSET are not the only genes located at 4p16 which are dysregulated by t(4;14)(p16;q32).*

2) *That a true t(4;14) target gene would be overexpressed or underexpressed at the mRNA level in all t(4;14) positive samples.*

C2.3 – Chapter 2 Results

Almost all of the results documented in this chapter are published in Blood in two manuscripts; Keats et al. 2003 and Keats et al. 2005[1,122]. Moreover, the overall patient outcome data includes an additional 36 months of follow up compared to the initial data presented in Keats et al. 2003[1], and the overall outcome associated with *FGFR3* expression and breakpoint type includes an additional 9 months of follow-up compared to the initial data presented in Keats et al. 2005[122].

C2.3.1 – The Expression of *FGFR3* in t(4;14) Positive Patients

The initially proposed mechanism by which t(4;14) contributed to myelomagenesis was the dysregulation of *FGFR3* on the der(14) by the 3' IgH regulatory regions (alpha enhancers)(Figure C2.1)[88]. As a result of the translocation the telomeric end of 4p16 is transferred to chromosome 14 to produce the der(14) chromosome. This brings *FGFR3*, which is not normally expressed in plasma cells or BMMC, into close proximity of the 3' IgH regulatory regions resulting in dysregulation. Since, *FGFR3* is not normally expressed in BMMC the dysregulation of *FGFR3* can be measured qualitatively with a standard RT-PCR reaction using *FGFR3* specific primers. As one of our initial goals was to determine if patients were expressing wild-type or mutated forms of *FGFR3* we amplified *FGFR3* transcripts from the BMMC samples of t(4;14) positive patients with two different sets of *FGFR3* primers; FGFR3-A and FGFR3-B. The primer sets amplify the transmembrane and tyrosine kinase II domains of *FGFR3*, respectively. These amplicons were chosen as they contain the majority of the known *FGFR3* activating mutations.

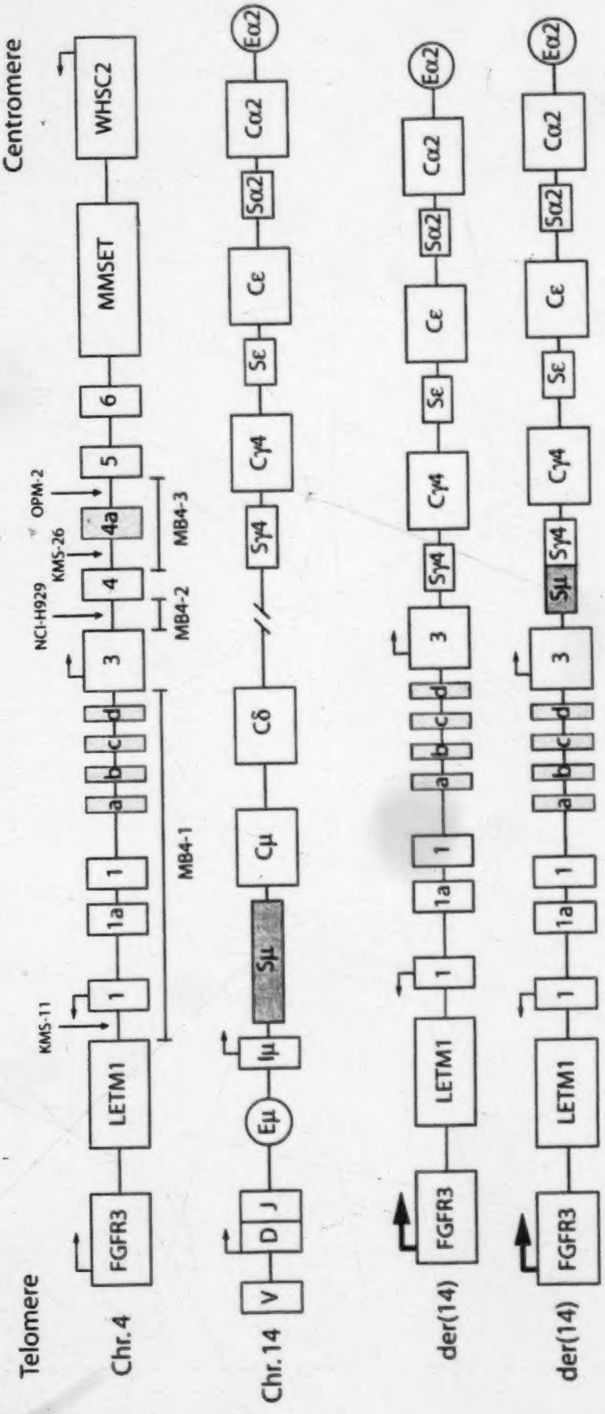

A)
Telomere
Centromere
Chr. 4
FGFR3
LETM1
KMS-11
1
1a
1
a
b
c
d
3
NCI-H929
4
KMS-26
4a
OPM-2
5
6
MMSET
WHSC2
MB4-1
MB4-2
MB4-3
Chr. 14
V
D
J
Eμ
Iμ
Sμ
Cμ
Cδ
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
B)
der(14)
FGFR3
LETM1
1
1a
1
a
b
c
d
3
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
der(14)
FGFR3
LETM1
1
1a
1
a
b
c
d
3
Sμ
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2

Figure C2.1 – The Mechanism of *FGFR3* Dysregulation on the der(14)

A) Descriptive outline of the chromosomal regions involved in t(4;14) (See Figure C1.2 for detailed description). B) Potential der(14) chromosomes created by t(4;14) showing the dysregulation of *FGFR3*, indicated by thick black arrows. The upper panel illustrates the typical der(14) with chromosome 4 rearranged directly into a downstream switch region. The lower panel illustrates the rare events with chromosome 4 jointed to switch mu and a subsequent class switch recombination event between switch mu and a downstream switch region, switch gamma4 in this illustration.

During our initial attempts to amplify *FGFR3* transcripts we were surprised to find several t(4;14) positive patients that did not express detectable levels of *FGFR3*. To determine if our results represented false-negatives we designed a third set of *FGFR3* specific primers, FGFR3-C. Furthermore, as we were using unpurified BMMC with variable levels of plasmacytosis we expanded the analysis to include all of the t(4;14) positive patients. Unexpectedly, only 31/44 (70.5%) t(4;14) positive patients had detectable levels of *FGFR3* expression (Figure C2.2) (Table C2.1). Analysis of sequential BM samples from 14 of the t(4;14) positive patients confirmed the *FGFR3* expression results. The 4 *FGFR3* non-expressors remained negative at all timepoints and the 10 *FGFR3* expressors remained positive at all timepoints. Moreover, the level of bone marrow plasmacytosis did not appear to influence the detection of *FGFR3* expression as the median bone marrow plasma cell percentage in the *FGFR3* non-expressors was 40% (range, 10-85%). Furthermore, *FGFR3* non-expressors are present in all three breakpoint sub-groups, so the expression of *FGFR3* appears to be independent of breakpoint type (Table C2.1).

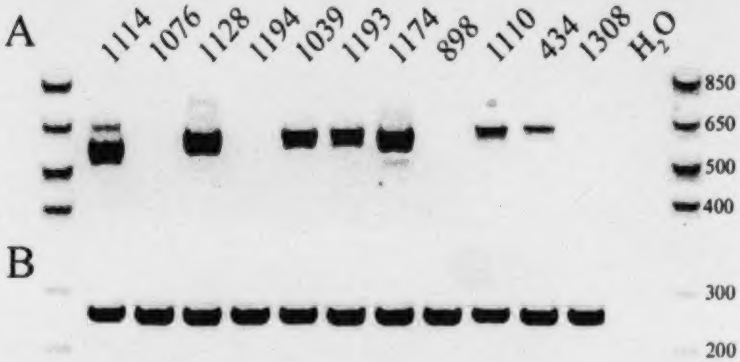


Figure C2.2 – *FGFR3* Transcripts are not Detectable in all t(4;14) Positive Patients

A) The qualitative expression of *FGFR3* was determined with the FGFR3-C primer set. A representative panel of eleven patients is shown of which only seven express detectable levels of *FGFR3*. The patients shown are the same patients shown in Figure C1.4 with patients;1114, 1076, 1128, 1194, 1039, and 1193 having MB4-1 breakpoints; 1174, 898, and 1110 having MB4-2 breakpoints; while 434 and 1308 have MB4-3 breakpoints. Of note patient 1308 is one of the rare patients with a predicted chromosome 14 breakpoint between JH6 and Imu as the JH-ms6r reaction is positive but the Iμ1-ms6r reaction is negative. B) The beta-2-microglobulin cDNA integrity test results are shown for each sample to confirm the quality of the cDNA template.

Table C2.1 – Frequency of *FGFR3* Expression in t(4;14) Positive Patients

Disease	Number of Patients with t(4;14)	Number of Patients Expressing *FGFR3*	Percent Expressing *FGFR3*
Myeloma	44	31	70.5%
MB4-1	31	23	74.2%
MB4-2	6	4	66.7%
MB4-3	7	4	57.1%
MGUS	2	2	100%
MB4-1	2	2	100%

FGFR3 expression was determined with the FGFR3-C primer set

C2.3.2 – Clinical Outcome Associated with t(4;14) and/or *FGFR3* Expression

The unexpected *FGFR3* expression results prompted us to reanalyze the survival profiles of the original patient cohort with respect to t(4;14) status and *FGFR3* expression. Within the original cohort of 208 myeloma patients with 31 t(4;14) positive patients we found 23 (74.2%) patients that expressed detectable levels of *FGFR3*. Since t(4;14) was associated with a poor prognosis, we expected the patients expressing the proposed t(4;14) target gene, FGFR3, would follow an aggressive clinical course. Moreover, since the FGFR3 non-expressing patients did not express the proposed target gene, we expected these patients would follow a clinical course similar to the t(4;14) negative cases. However, we did not observe a significant difference in the median survival between the t(4;14) positive patients when segregated based on their *FGFR3* expression (P=0.250)(Figure C2.3). The median survival was 813 days versus 695 days for *FGFR3* expressors and non-expressors respectively. Surprisingly, in this exploratory analysis only the *FGFR3* non-expressors appeared to have an inferior outcome relative to patients lacking the t(4;14) translocation (P=0.016 for *FGFR3* non-expressors and P=0.202 for *FGFR3* expressors compared to t(4;14) negative cases)(Figure C2.3).

Based on these results and the fact that during data collection Rasmussen et al. showed t(4;14), as defined by *FGFR3* expression, did not influence survival we screened the 177 t(4;14) negative patients for *FGFR3* expression[120]. None of the t(4;14) negative cases expressed detectable levels of *FGFR3*. When we analyzed the clinical outcome associated with *FGFR3* expression alone in the 23 *FGFR3* expressors compared to the 185 non-expressors we found no significant difference in survival, median survival of 813 days compared to 1242 days, respectively (P=0.251)(Figure C2.3). Given the results when patients within this cohort are grouped based on t(4;14) and *FGFR3* expression this is an expected result.

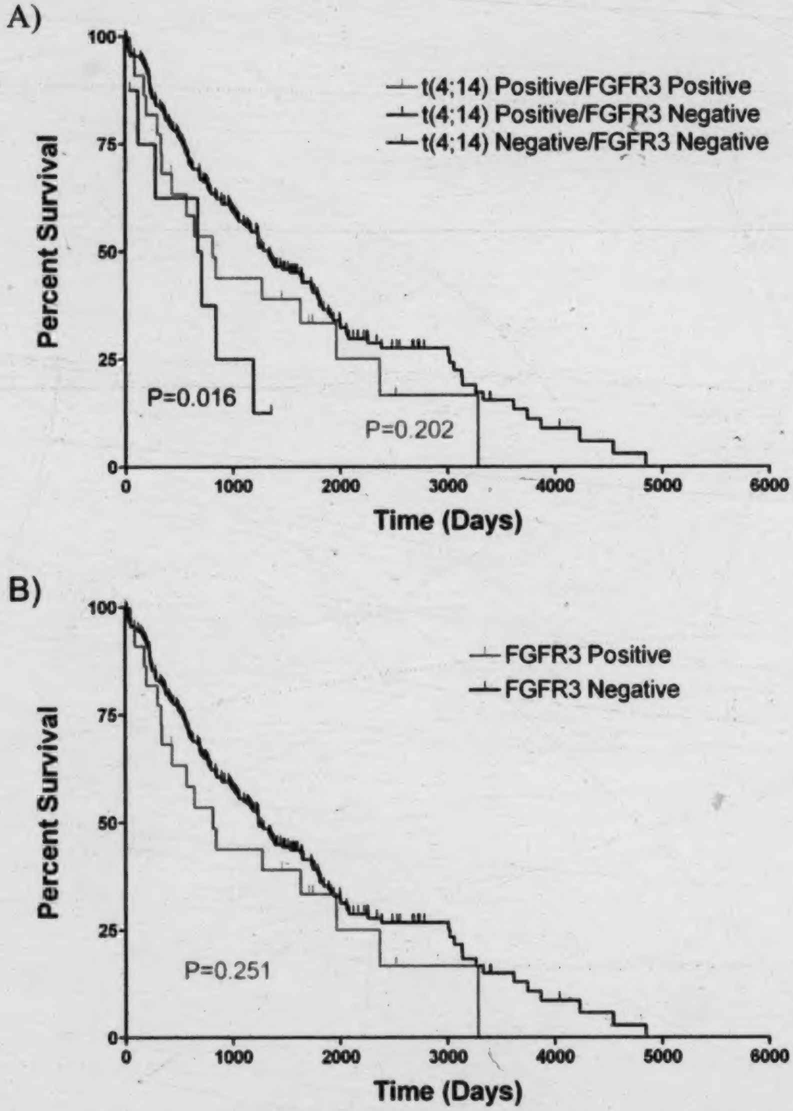


Figure C2.3 – The Influence of t(4;14) and *FGFR3* Expression on Overall Survival

A) Kaplan-Meier survival plot comparing the survival of the 23 t(4;14) positive/FGFR3 positive patients, the 8 t(4;14) positive/FGFR3 negative patients, and the 177 t(4;14) negative/FGFR3 negative patients. No significant difference exists between the t(4;14) positive/FGFR3 positive and t(4;14) positive/FGFR3 negative patients; P=0.250, median survival 813 days versus 692 days. No significant difference exists between the t(4;14) positive/FGFR3 positive and t(4;14) negative/FGFR3 negative patients; P=0.202, median survival 813 days versus 1338 days; HR 1.39, 95% CI 0.82-2.60. A significant difference exists between the t(4;14) positive/FGFR3 negative and t(4;14) negative/FGFR3 negative patients; P=0.016, median survival 692 days versus 1338 days; HR 2.45; 95% CI 1.30-13.23. B) Kaplan-Meier survival plot comparing the survival of 23 *FGFR3* expressing and 185 *FGFR3* non-expressing myeloma patients. No significant difference exists between these two groups; P=0.251, median survival 813 days versus 1242 days, HR 1.34; 95% CI 0.79-2.47.

C2.3.3 – The der(14) Chromosome is Undetectable in *FGFR3* non-expressers

The lack of *FGFR3* expression may reflect a loss of the der(14) chromosome created by the translocation or an inhibition of *FGFR3* expression. To investigate whether the der(14) chromosome was detectable in our t(4;14) positive patients we screened these patients for detectable der(14) MMSET-IgH hybrid transcripts (Figure C2.4)(Also Figure C1.1). Because of the large number of potential der(14) MMSET-IgH transcripts we screened each t(4;14) positive BM sample with eight different RT-PCR reactions. Half the reactions utilized a 5' *MMSET* exon 1 (5' ME1) primer in combination with consensus 3' primers for the CH1 exon of IgM or IgD or IgG or IgA, while the other half utilized a 5' *MMSET* exon 3 (5' ME3) primer with the same IgH primers. The *MMSET* exon 1 reactions will only detect a der(14) MMSET-IgH hybrid transcript if the chromosome 4 breakpoint is downstream of *MMSET* exon 1 (Figure C2.4). Moreover, reactions using a 5' ME3 primer will only amplify der(14) MMSET-IgH hybrid transcript products from MB4-2 and MB4-3 patients (Figure C2.4).

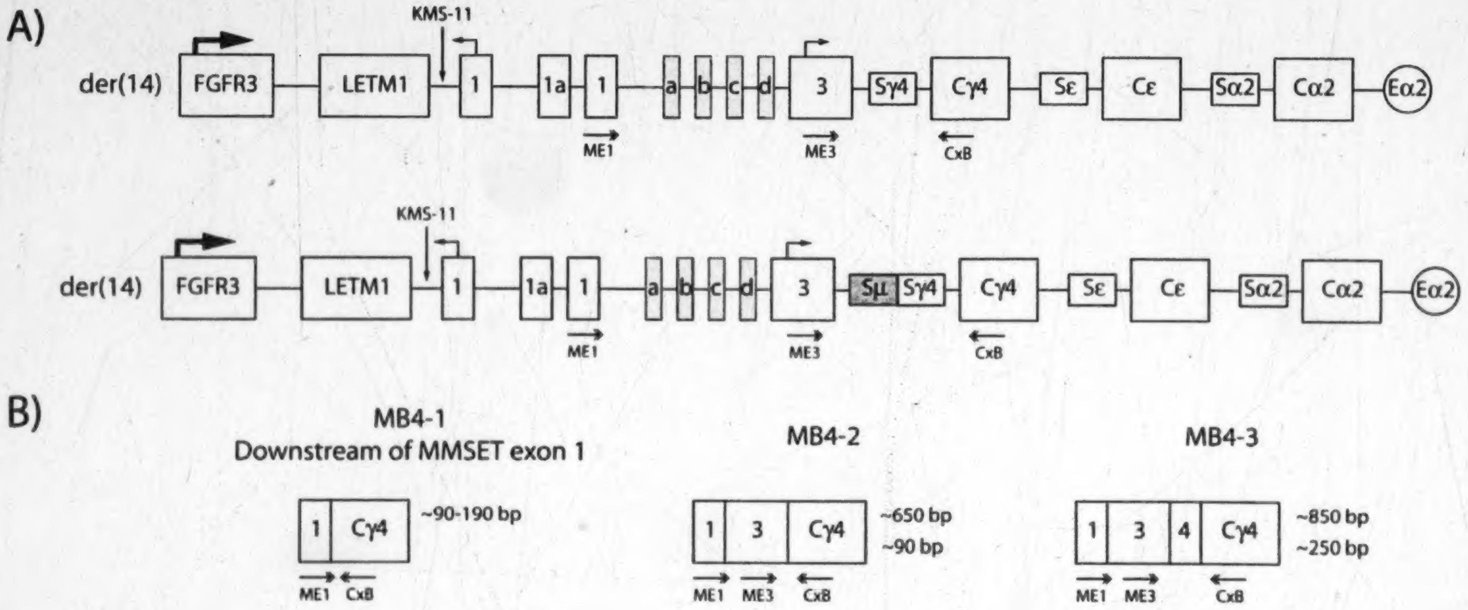

A)
der(14)
FGFR3
LETM1
KMS-11
1
1a
1
a
b
c
d
3
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
ME1
ME3
CxB
der(14)
FGFR3
LETM1
KMS-11
1
1a
1
a
b
c
d
3
Sμ
Sγ4
Cγ4
Sε
Cε
Sα2
Cα2
Eα2
ME1
ME3
CxB
B)
MB4-1
Downstream of MMSET exon 1
1
Cγ4
~90-190 bp
ME1
CxB
MB4-2
1
3
Cγ4
~650 bp
~90 bp
ME1
ME3
CxB
MB4-3
1
3
4
Cγ4
~850 bp
~250 bp
ME1
ME3
CxB

Figure C2.4 – Description of der(14) Screening Assays

A) The typical der(14) chromosomes created by t(4;14) showing the location of the various primers used to detect der(14) MMSET-IgH hybrid transcripts. Note the location of *MMSET* exon 1 containing the most telomeric primer, 5' ME1, and the characterized KMS-11 breakpoint upstream of this exon between *LETM1* exons 1 and 2. B) The predicted mRNA species detected by the der(14) MMSET-IgH hybrid transcript assays for each breakpoint type are shown. The observed product sizes for each reaction are shown to the right of each predicted mRNA species with the ME1 reaction product sizes above and the ME3 reaction product sizes below. The 3' primers specific to each heavy chain isotype are noted as CxB to reflect the four different primers used; CmuB, CdeltaB, CgammaB, and CalphaB

In our t(4;14) positive patient population we detected der(14) MMSET-IgH hybrid transcripts in 11/44 (25%) of the myeloma patients and in one of the MGUS patients (Table C2.2). However, this number is an under representation as the majority of our patients have the MB4-1 breakpoint and some false-negatives are expected within this group. Moreover, a detailed mapping of the published genomic breakpoints from MB4-1 patients showed 11/15 (73%) are upstream of *MMSET* exon 1 and thus would not generate a product in the der(14) MMSET-IgH hybrid transcript assays (See Chapter 4). Therefore, assuming the published breakpoint sequences reflect the natural distribution in myeloma patients our observation that 21.7% of the MB4-1 patients with *FGFR3* expression have detectable der(14) MMSET-IgH hybrid transcripts is quite reasonable (Table C2.2). Since the breakpoints in MB4-2 and MB4-3 t(4;14) positive patients are predicted to be downstream of *MMSET* exons 1 and 3, we expected these patients to have detectable der(14) MMSET-IgH hybrid transcripts. In the 6 MB4-2 and 7 MB4-3 patients; 6/8 *FGFR3* expressors had detectable der(14) products in both reactions while 5/5 *FGFR3* non-expressors lacked detectable der(14) hybrid transcripts. Similarly, none of the MB4-1 *FGFR3* non-expressors had a detectable der(14) product. Since none of the *FGFR3* non-expressors have detectable der(14) MMSET-IgH hybrid transcripts, we believe the lack of *FGFR3* expression in a subset of t(4;14) positive myeloma patients results from the loss of the der(14) chromosome created by the translocation. Why 25% of our MB4-2 and MB4-3 patients with detectable levels of *FGFR3* expression do not have detectable der(14) MMSET-IgH hybrid transcripts is unknown. However, two of the myeloma cell lines, NCI-H929 (MB4-2) and OPM-2 (MB4-3) also fail to generate a product in these reactions. In both cases the genomic der(14) was cloned by ourselves or others and no mechanistic reason for this phenomenon has been determined (See Chapter 4).

Table C2.2 – Results of the der(14) MMSET-IgH Hybrid Transcript Assays

Disease	Detection of der(14) MMSET-IgH Hybrid Transcripts in t(4;14) Positive Samples		
	All Patients	**FGFR3 Positives**	**FGFR3 Negatives**
Myeloma	11/44 (25%)	11/23 (47.8%)	0/8 (0%)
MB4-1	5/31 (16.1%)	5/23 (21.7%)	0/8 (0%)
MB4-2	2/6 (33.3%)	2/4 (50%)	0/2 (0%)
MB4-3	4/7 (57.1%)	4/4 (100%)	0/3 (0%)
MGUS	1/2	1/2	-
MB4-1	1/2	1/2	-

Interestingly, for 9 of the 11 patients, their der(14) MMSET-IgH hybrid transcript products were detected with CH1 primers specific to their clinical isotype, IgG or IgA in all cases (Table C2.3). Therefore, in the majority of the cases the class switch recombination event occurring on the functionally rearranged chromosome 14 also occurs on the chromosome 14 involved in the translocation event. It is not known if the two patients with der(14) hybrid transcripts linked to non-clinical isotypes reflects a mis-targeting of the class switch recombination machinery or secondary events.

Table C2.3 – Correlation between Clinical Isotype and der(14) Result

Patient	Breakpoint Type	der(14) Associated Heavy Chain	Clinical Isotype	Sequenced Isotype
773	MB4-1	**IgG**	**IgA-Kappa**	ND
945	MB4-1	IgG	IgG-Kappa	ND
1237	MB4-1	IgG	IgG-Kappa	IgG
1183	MB4-1	IgG	IgG-Lambda	ND
1091	MB4-1	IgG	IgG-Kappa	IgG
1174*	MB4-2	IgG	IgG-Lambda	IgG
1110	MB4-2	IgA	IgA-Kappa	ND
434*	MB4-3	IgG	IgG-Lambda	IgG
1661*	MB4-3	IgG	IgG-Kappa	ND
1394*	MB4-3	**IgA**	**IgG-Kappa**	IgG
657	MB4-3	IgA	IgA-Kappa	ND

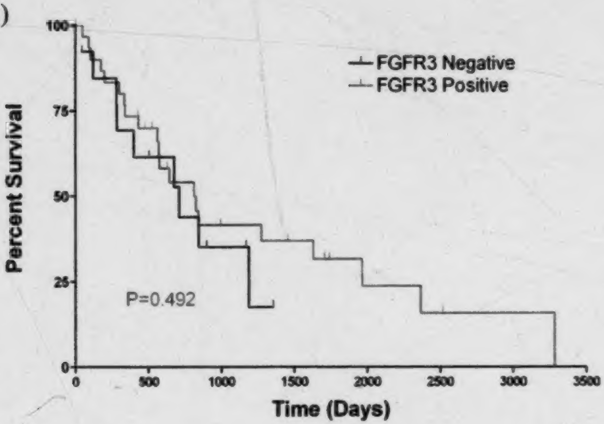
Patients with der(14) MMSET-IgH hybrid transcripts detected with a primer specific to a heavy chain isotype different from their clinical isotype are noted in bold font. The asterisks indicates patients where the der(14) genomic breakpoint was cloned and thus confirmed the der(14) MMSET-IgH hybrid transcript results (See Chapter 4). Not Done, ND.

C2.3.4 – Exploratory Clinical Analysis

Though the poor prognosis associated with t(4;14) is now well accepted, several issues require clarification. First, the impact of *FGFR3* expression on the survival of t(4;14) positive patients is unclear. When t(4;14) positive patients from our original cohort are separated into *FGFR3* expressors and non-expressors it is only in the non-expressors that the poor prognosis remains statistically significant. One potential explanation for this phenomenon is that the apparent loss of der(14) in the *FGFR3* non-expressors may in fact be the poor prognostic variable[43]. Alternatively, our original patient cohort may not be of sufficient size to determine the clinical effect of *FGFR3* expression. In an attempt to address this issue we analyzed the clinical outcome associated with *FGFR3* expression in the t(4;14) positive patients from our expanded cohort (Figure C2.5). This expanded cohort includes 44 t(4;14) positive patients with 31 *FGFR3* expressors and 13 *FGFR3* non-expressors. In this exploratory analysis, no statistically significant difference exists between *FGFR3* expressors and non-expressors with a median survival of 709 and 813 days, respectively (P=0.492; HR=1.30, 95% CI 0.57-3.18)(Figure C2.5). This suggests the poor outcome associated with t(4;14) is directly associated with the translocation and not *FGFR3* expression or the potential loss of the der(14) chromosome in *FGFR3* non-expressors.

Therefore, the second proposed t(4;14) target gene, *MMSET*, appears to be a strong candidate gene. However, the der(4) IgH-MMSET hybrid transcript assays identify a critical issue with this gene. The predicted breakpoint locations defined by the hybrid transcript assays identify patients capable of overexpressing wild-type or truncated MMSET protein variants. In patients with MB4-1 breakpoints the breakpoint is upstream of the proper translation initiation site in MMSET exon 3, and thus these patients are predicted to overexpress wild-type MMSET protein variants. However, patients with MB4-2 and MB4-3 breakpoints have genomic breakpoints downstream of MMSET exon 3, and thus the proper translation initiation site is lost. Therefore, these patients are predicted to overexpress truncated MMSET protein variants. To determine if the ability to encode a full length MMSET protein influenced overall survival, we compared the survival difference between patients with MB4-1 breakpoints (wild-type MMSET producers) versus patients with MB4-2 or MB4-3 breakpoints (truncated MMSET producers). The 44 t(4;14) positive patients from the expanded cohort were analyzed in this analysis comparing the clinical outcome of the 31 MB4-1 patients versus the 13 patients with either MB4-2 or MB4-3 breakpoints. No significant difference in clinical outcome was associated with the ability to produce or not produce wild-type MMSET as the median survival was 813 and 575 days, respectively (P=0.537, HR=0.80, 95% CI 0.35-1.74) (Figure C2.5). Therefore, the poor prognosis associated with t(4;14) appears to be independent of *FGFR3* expression and breakpoint type defined by the der(4) IgH-MMSET hybrid transcript assays.

A)



B)

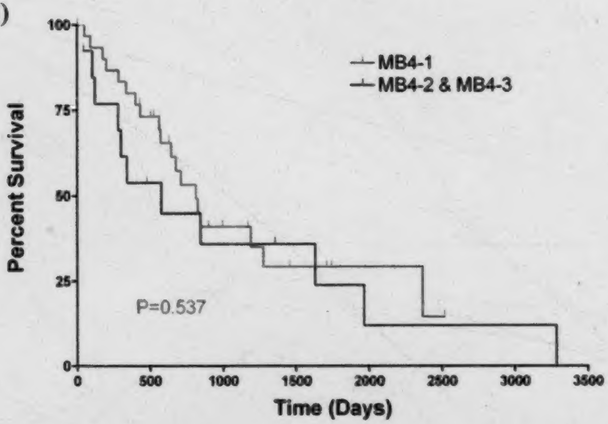


Figure C2.5 – The Poor Outcome Associated with t(4;14) is Independent of FGFR3 Expression and Breakpoint Type

A) Kaplan-Meier survival curve comparing 31 *FGFR3* expressing and 13 *FGFR3* non-expressing t(4;14) positive patients with a median survival of 813 and 709 days, respectively (P=0.492; HR=1.3, 95% CI 0.57-3.18). B) Kaplan-Meier survival curve comparing 31 MB4-1 patients and 13 patients with either MB4-2 or MB4-3 breakpoints with a median survival of 813 and 575 days, respectively (P=0.537; HR=0.80, 95% CI 0.35-1.72).

C2.3.3 – Alternative Target Genes

Based on our survival analysis and the lack of detectable *FGFR3* expression in approximately 30% of t(4;14) positive myeloma patients, *FGFR3* is unlikely to be the major target gene of the t(4;14) translocation. Furthermore, t(4;14) appears to be associated with a poor outcome irrespective of *FGFR3* expression or breakpoint type. Therefore, we hypothesized that universal features must exist in these patients, which promote myelomagenesis and an aggressive clinical course. We assumed a central universal feature would be the dysregulation of a target gene on chromosome 4, as very few non-immunoglobulin genes exist within the breakpoint region on chromosome 14. Our working hypothesis was that a t(4;14) target gene would be overexpressed or underexpressed at the mRNA level in all t(4;14) positive samples. As the effects of the mu enhancer and 3' IgH regulatory regions may act over large distances, we identified all known genes from 4p16 within approximately 500 kb of either side of the breakpoint sites on chromosome 4 (Figure C2.6). An annotated human genome contig of 1036 kb from the most recent human genome build, Build 35.1, was used to identify potential target genes flanking the t(4;14) breakpoints. This region of chromosome 4 contains 17 genes including several previously proposed t(4;14) target genes, such as *TACC3*, *FGFR3*, *LETM1*, *MMSET*, and *WHSC2* (Figure C2.6).

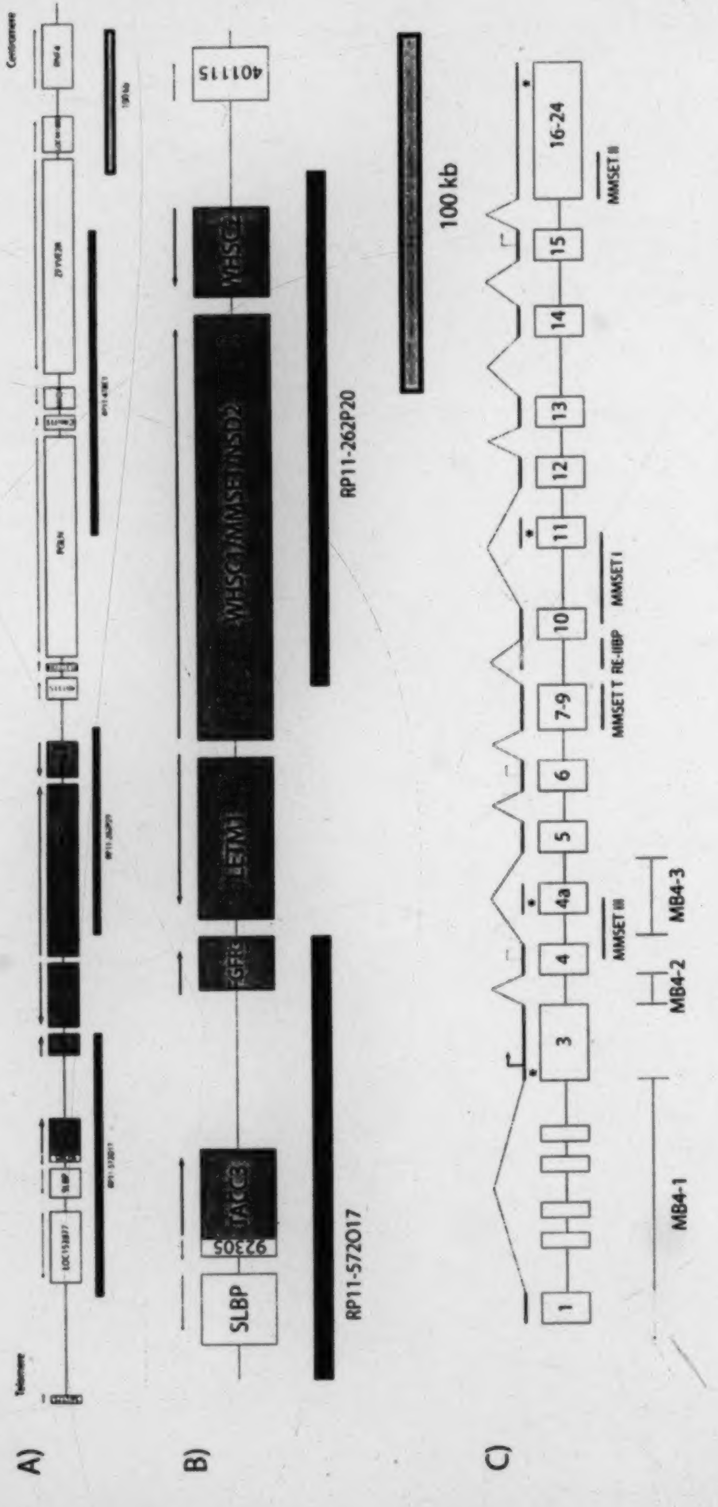

A)
Telomere
Centromere
SLBP
POLN
ZFYVE28
100 kb
B)
SLBP
92305
TACC3
FGFR3
LETM1
WHSC1/MMSET/NSD2
WHSC2
401115
RP11-572O17
RP11-262P20
100 kb
C)
1
3
4
4a
5
6
7-9
10
11
12
13
14
15
16-24
MMSET III
MMSET T
RE-IIBP
MMSET I
MMSET II
MB4-1
MB4-2
MB4-3

Figure C2.6 – Genes Flanking the t(4;14) Breakpoints on 4p16

A) The entire 1036 kb of contig NT_037623.4 from human genome project build 35.1 is shown to scale. The previously proposed alternative t(4;14) target genes; *TACC3*, *FGFR3*, *LETM1*, *MMSET*, and *WHSC2* are noted by grey boxes. Directional arrows indicate the direction of transcription for each gene. B) A zoomed region of contig NT_037623.4 containing the previously proposed t(4;14) target genes. C) Exon-intron structure diagram of the *MMSET* gene. The breakpoint cluster regions for MB4-1, MB4-2, and MB4-3 are shown. The mRNA species encoding the full length wild-type MMSET II protein is produced by the splicing pattern indicated by solid lines. Alternative splicing events producing transcripts which encode MMSET III and MMSET I are indicated by dotted grey lines. In-frame stop codons are indicated by an asterisk. The proper MMSET translation initiation site is indicated by a solid black arrow, while the proposed alternative translation initiation sites in exon 4 and 6 identified by Chesi et al. are indicated by grey arrows[91]. The point of transcription initiation for the transcript encoding RE-IIBP is indicated by grey square boxes and the translation initiation site is indicated by a dotted black arrow. The approximate locations of the individual qRT-PCR reactions are indicated by dotted black lines. The MMSET Total (T) reaction spans exons 8 and 9, while MMSET II spans exons 16 and 17.

We initiated the study with all previously proposed target genes with the intention of expanding our analysis if our most distant telomeric or centromeric genes fit the hypothesis. Preliminary experiments with qualitative RT-PCR confirmed our suspicion that many of the proposed t(4;14) target genes other than *FGFR3* were expressed at detectable levels in both t(4;14) positive and negative cell lines and normal bone marrow. Therefore, the study was conducted on a panel of cell lines and CD138 purified patient plasma cells using quantitative RT-PCR with Taqman probes specific to each potential target gene.

An initial pilot experiment was conducted in the cell line panel to determine the expression of *TACC3*, *FGFR3*, *LETM1*, *MMSET* total (MMSET type I and II encoding transcripts), the alternative *MMSET* transcription event encoding RE-IIBP, and *WHSC2*. The panel of cell lines included the Burkitt's lymphoma cell lines Raji and Daudi as reference cell lines; the t(4;14) negative myeloma cell lines RPMI-8226, U266, KMS-12BM, and KMS12PE; and the t(4;14) positive myeloma cell lines KMS-11, NCI-H929, KMS-18, JIM3, OPM-2, and LP-1. Within this small panel of myeloma cell lines *TACC3*, *FGFR3*, *LETM1*, *MMSET* total, and RE-IIBP transcripts were dysregulated in at least one of the t(4;14) positive cell lines (Table C2.4). However, the only transcripts expressed at higher levels in the t(4;14) positive cell lines compared to the t(4;14) negative cell lines are *MMSET* total (transcripts encoding MMSET I, II, and III) and RE-IIBP transcripts (Table C2.4). As *TACC3* and *WHSC2*, the most distant telomeric and centromeric gene respectively, did not fit our working hypothesis; we moved the study into patient samples and did not expand our analysis to other genes.

Table C2.4 - Quantitative Relative Expression Level of Genes Located at 4p16.3 in Cell Lines

Cell Line	TACC3	FGFR3	LETM1	MMSET Total	RE-IIBP	WHSC2
Daudi	1.51	0.00	1.02	1.02	1.44	1.41
Raji	1.00	0.00	1.00	1.00	1.00	1.00
RPMI-8226	1.51	0.00	1.26	1.11	1.65	1.17
U266	1.16	0.00	1.23	1.21	1.98	0.94
KMS-12BM	0.81	0.16	1.01	0.74	0.76	0.98
KMS-12PE	1.20	0.21	1.63	1.18	1.51	1.42
KMS-11 MB4-1, POS	1.01	87.86	0.36	2.13	4.27	0.51
NCI-H929 MB4-2, POS	3.61	4.23	1.04	2.16	4.85	1.03
KMS-18 MB4-2, POS	0.79	100.00	6.44	4.65	13.83	0.47
JIM3 MB4-2, NEG	0.33	0.02	1.12	7.87	14.45	0.39
OPM-2 MB4-3, POS	1.82	64.47	0.54	5.39	12.82	1.13
LP-1 MB4-2*, POS	1.52	0.80	5.19	1.47	2.32	1.81

The relative expression level (REL) for most reactions is a comparison with the expression level detected in Raji, which is used as the normalizer with an expression level of 1. The only exception is the REL for FGFR3, which uses the expression level detected in KMS-18 as the normalizer with an expression level of 100. The panel of cell lines includes two Burkitt's Lymphoma lines (Daudi, Raji), four t(4;14) negative myeloma cell lines (RPMI-8226, U266, KMS-12-BM, KMS-12-PE), and six t(4;14) positive myeloma cell lines (KMS-11, NCI-H929, KMS-18, JIM3, OPM-2, LP-1). The breakpoint type and FGFR3 expression status of each t(4;14) positive cell line is noted. The asterisk notes a unique der(4) breakpoint occuring in a downstream switch region that did not rearrange with switch mu, which increases the distance between the mu enhancer and genes from chromosome 4.

The cell line pilot experiments identified transcripts originating from the *MMSET* gene as being universally dysregulated in t(4;14) positive lines. As a number of alternative splicing events and transcription events occur within the *MMSET* gene, we expanded our analysis to assay the different transcripts encoding the four principle protein variants of *MMSET* (i.e. Exon 4a/MMSET III, MMSET I, MMSET II, and RE-IIBP)(Figure C2.6). The expression of potential target genes in a patient population was determined from CD138 positive plasma cells isolated from diagnostic or relapse bone marrow samples of 17 myeloma patients, of which 6 are t(4;14) positive and 11 are t(4;14) negative. We purified the diagnostic and relapse sample from one t(4;14) positive patient, bringing the total number of purified bone marrow samples to 18. The 6 t(4;14) positive patients included equal numbers of *FGFR3* expressors and non-expressors. Within this series of patient samples only transcripts originating from the *MMSET* locus; *MMSET* total, MMSET III, MMSET I, MMSET II, and RE-IIBP were significantly dysregulated in the t(4;14) positive patients (Table C2.5). The dysregulation of *FGFR3* and *TACC3* approached statistical significance. However, for *FGFR3* it did not reach statistical significance as 50% of the t(4;14) positive patients in this series are *FGFR3* non-expressors. *TACC3* also approached statistical significance but this is largely due to a single high expressor. Of interest this t(4;14) positive patient lacked *FGFR3* expression but had an 18 fold increase in the expression of *TACC3*. This may reflect a translocation event where *FGFR3* is deleted from der(14) but *TACC3* is maintained and subsequently overexpressed. Therefore, within the panel of genes analyzed, only transcripts originating from the *MMSET* locus fit our working hypothesis, suggesting that the target gene of this translocation is likely the protein product of one of these transcripts. Interestingly, the expression level of RE-IIBP and other *MMSET* transcripts in *ex vivo* cells considerably exceeds the level detected in myeloma cell lines.

Table C2.5 – Quantitative Relative Expression of Potential t(4;14) Target Genes in *ex vivo* Plasma Cells

Patient	TACC3	FGFR3	LETM1	MMSET Total	MMSET III	MMSET I	MMSET II	RE-IIBP	WHSC2
1335	1.34	0.56	8.36	4.66	10.08	11.13	3.24	11.39	11.06
1336	0.32	0.04	1.84	0.59	2.40	0.89	0.40	0.82	0.97
1677*	0.33	0.04	1.54	1.04	2.17	2.49	0.90	0.93	0.95
1676*	0.61	0.05	2.56	0.92	2.09	1.84	1.13	2.31	2.19
1668*	0.19	0.01	1.19	0.61	0.96	0.94	0.68	0.29	1.03
1644*	1.80	0.55	6.66	17.6	5.18	1.54	2.91	6.74	6.41
1645*	5.37	8.25	5.69	1.82	2.76	0.81	1.40	3.47	4.51
1364	0.95	0.05	4.04	1.72	2.08	2.27	1.39	2.92	2.49
1243	0.55	0.05	3.53	1.14	2.83	0.71	0.85	1.53	2.72
527	1.03	0.01	1.20	0.71	1.63	0.68	0.55	0.53	1.30
1660*	0.42	0.01	1.56	0.99	1.93	1.09	0.73	1.08	1.18
1504 MB4-1, NEG	18.78	1.22	0.48	77.81	77.29	167.65	29.73	286.73	0.89
1560* MB4-1, NEG	0.76	0.63	1.65	24.53	71.95	72.58	17.21	34.77	1.25
1649* MB4-1, POS	0.74	339.10	4.49	22.13	95.90	14.02	22.84	140.61	2.97
1237 MB4-1, POS	2.54	931.97	11.48	71.59	158.23	42.47	43.93	156.32	4.56
1308 MB4-3, NEG	0.23	0.07	1.58	45.25	1.85	70.20	34.78	124.79	2.33
1661 MB4-3, POS	0.21	58.10	0.90	12.47	0.65	27.86	9.60	24.48	0.79
1661* MB4-3, POS	4.02	139.80	3.50	40.50	1.40	30.13	31.41	38.23	2.14
t(4;14) Negative Mean REL (Mean ΔCt)	1.17 (8.63)	0.87 (12.6)	3.47 (4.91)	2.89 (7.49)	3.10 (10.4)	2.22 (9.81)	1.29 (4.78)	2.91 (10.1)	3.16 (5.67)
t(4;14) Positive Mean REL (Mean ΔCt)	3.90 (7.84)	210 (4.90)	3.44 (5.28)	42.0 (2.91)	58.2 (7.82)	60.7 (4.84)	27.1 (0.23)	115 (4.48)	2.13 (5.99)
t-Test	0.182	0.054	0.904	<0.001	0.005	0.001	<0.001	0.001	0.497

The relative expression level was determined by the ΔΔCt analysis method using the Raji cell line as a reference expression level of 1 for all reactions except FGFR3. Since Raji does not express *FGFR3* we used the KMS-18 myeloma cell line as a reference expression level of 100 as this t(4;14) positive cell line had the highest expression level. Diagnostic patient samples are indicated by asterisks. For t(4;14) positive samples the breakpoint type and *FGFR3* expression status are noted below the patient identifier. The mean ΔCt (Ct^{Target}-Ct^{GAPDH}) is shown to indicate the differences in absolute expression levels between each gene (The lower the ΔCt value the greater the expression level).

C2.3.6 – RE-IIBP is Overexpressed in BMMC of t(4;14) Positive Patients

The apparent loss of der(14) in the *FGFR3* non-expressors suggests the true t(4;14) target gene(s) is present on the der(4) chromosome. In the small panel of purified patient plasma cells the only transcripts that are universally dysregulated in t(4;14) positive samples originate from the *MMSET* gene. However, the transcripts corresponding to MMSET I and II will not produce the same proteins in all patients as patients with MB4-2 or MB4-3 breakpoints lose the first or first and second translated exons of *MMSET*, respectively. Therefore, the only universally dysregulated transcript producing an identical protein product in all t(4;14) positive patients is RE-IIBP. Since RE-IIBP represents an overexpressed transcript that does not vary between breakpoint types, we confirmed its overexpression on a larger panel of unpurified myeloma BMMCs. This group of patients included 25 t(4;14) negative and 21 t(4;14) positive patients. Only t(4;14) positive patients with bone marrow plasma cell percentages above 35% were included in the study. The median relative expression level of RE-IIBP in the t(4;14) negative patients was 2.80 (range, 0.42-7.91), while in the t(4;14) positive samples it was 90.59 (range, 3.86-366.67), P<0.001 (Figure C2.7). The t(4;14) positive patient with the lowest expression, 1308, was also tested as a purified sample and had an expression level of 124.78. Thus the low level is likely due to a low plasma cell percentage within the unpurified sample. Therefore, the dysregulation of RE-IIBP encoding transcripts appears to be a universal event of t(4;14) positive multiple myeloma.

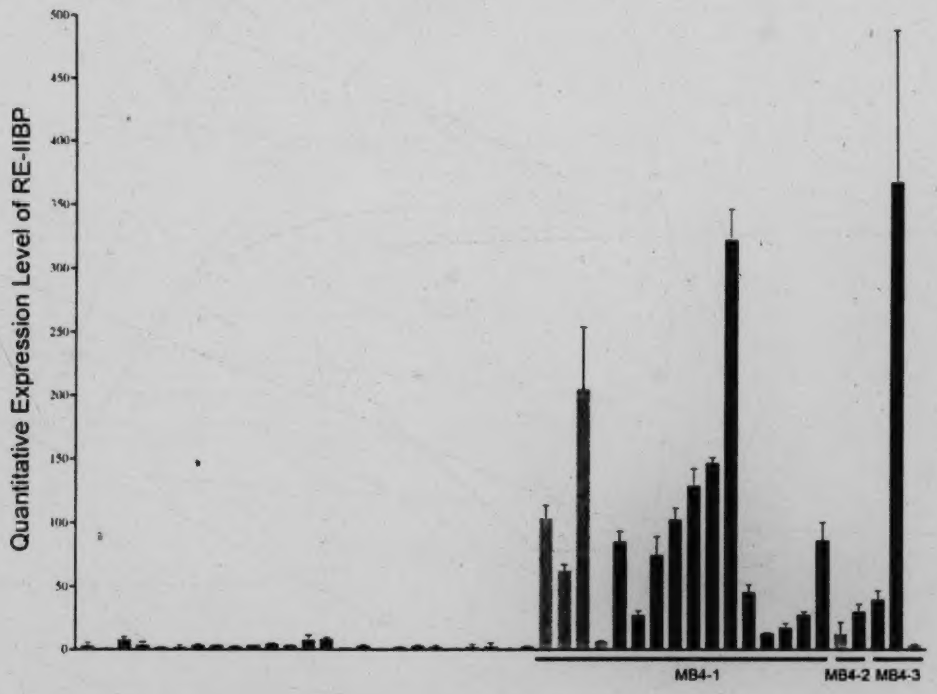


Figure C2.7 – RE-IIBP Expression in BMMC

The quantitative expression level of RE-IIBP encoding transcripts was determined in 25 t(4;14) negative and 21 t(4;14) positive patients. The t(4;14) negative patients were matched to the t(4;14) positive patients based on percentage of bone marrow plasma cells, sex, age at diagnosis, and clinical isotype. The t(4;14) positive patients are shown on the right and black underlines indicated each patients breakpoint type. The RE-IIBP expression levels in the t(4;14) positive/FGFR3 negative patients are indicated by grey bars.

C2.4.1 – Chapter Conclusions

The principle goal of this chapter was to identify potential t(4;14)(p16;q32) target genes. The initially proposed target gene of this translocation was *FGFR3*[88], however, several other potential target genes have been suggested since the original discovery of this translocation[1,91,92]. Therefore, we initiated a study to identify the genes dysregulated by the translocation. Ultimately, should a true target gene be identified and functionally validated, it would likely provide an excellent therapeutic target in this subset of myeloma patients with a poor overall outcome.

Experiments investigating our initial hypothesis that FGFR3 and MMSET would not be the only genes located at 4p16 which are dysregulated by t(4;14)(p16;q32) produced several unexpected yet fundamental results. First, the initially proposed target gene, *FGFR3*, was not expressed in approximately 30% of the t(4;14) positive patients at diagnosis. To our surprise, both t(4;14) positive/FGFR3 positive and t(4;14) positive/FGFR3 negative patients have a poor clinical outcome. Since these two groups have similar outcomes, t(4;14) appears to be a poor prognostic indicator irrespective of *FGFR3*. Interestingly, our analysis of sequential bone marrow samples did not identify any changes in *FGFR3* expression, suggesting the der(14) is selectively maintained in FGFR3 expressing t(4;14) positive patients. This could reflect a requirement for FGFR3 overexpression in this subset of patients or a gene dosage requirement associated with other genes on der(14). Our attempt to determine why *FGFR3* is not overexpressed in approximately 30% of the t(4;14) positive patients resulted in a paradigm shift in our understanding of IgH translocations in multiple myeloma. Using a series of hybrid transcript assays specific for the der(14) we were able to show that the der(14) chromosome, which should harbor the overexpressed *FGFR3* allele, was undetectable in all *FGFR3* non-expressing patients, suggesting the der(14) is lost in 30% of t(4;14) positive patients. The implications of this observation are dramatic. First, it suggests the IgH translocations in myeloma are not always reciprocal as held by the current dogma. Second, it calls into question many of the FISH studies looking at this translocation, as most assays are designed to detect the der(14) chromosome and thus the numbers represented in these studies are likely under representations of the actual frequency of t(4;14) in myeloma patients. Furthermore, the der(14) MMSET-IgH hybrid transcript assays showed that in all cases, an apparent class switch recombination event had occurred on the chromosome 14 involved in the translocation. Interestingly, in most cases the apparent class switch recombination involved the same IgH isotype produced by each specific patient.

Based on these initial observations we examined the expression of all proposed t(4;14) target genes using quantitative RT-PCR. Several of the investigated genes were dysregulated in individual t(4;14) positive patients and cell lines proving our initial hypothesis that *FGFR3* and *MMSET* are not the only genes located at 4p16 that are dysregulated by t(4;14). However, only transcripts originating from the *MMSET* gene fit our second hypothesis that a true t(4;14) target gene would be overexpressed or underexpressed at the mRNA level in all t(4;14) positive samples. Three different transcripts encoding the open reading frames of MMSET I, MMSET II, and RE-IIBP were universally overexpressed in all t(4;14) positive patients. Therefore, the second proposed t(4;14) target gene, *MMSET*, appears to be the true target gene. However, the three different transcripts are predicted to result in the overexpression of three different

MMSET protein variants. Moreover, not all patients are capable of overexpressing wild-type MMSET I and MMSET II protein variants as the genomic breakpoint is downstream of the proper translation initiation site in approximately 30% of the t(4;14) positive patients (MB4-2 and MB4-3 patients). Therefore, it remains to be determined which transcript or transcript(s) contribute to myelomagenesis. Furthermore, since the clinical outcome is not significantly different between patients with the ability to produce wild-type (MB4-1) versus truncated (MB4-2 and MB4-3) versions of MMSET proteins it remains to be determined if the transcripts encoding truncated versions of MMSET produce a protein products and if produced whether they maintain a functional capability.

Chapter 3

Analysis of MMSET Protein Variants Associated with t(4;14) Myeloma

C3.1.1 - Brief Introduction

By quantitative RT-PCR the only universally dysregulated transcripts in t(4;14) positive myeloma, from 4p16, are from the *MMSET* locus[122]. However, several transcripts from this locus are dysregulated and they all encoded different protein variants. The different transcripts originate from two principle transcription initiation events and several alternative splicing events. The first transcription initiation site is upstream of *MMSET* exon 3, which contains the proper translation initiation site[90,91]. Alternative splicing events within transcripts originating upstream of *MMSET* exon 3 produce three different mRNA species. The two principle protein variants MMSET I and MMSET II are encoded by mRNAs including or skipping *MMSET* exon 11, respectively[91]. A third protein variant, MMSET III, is encoded by mRNAs containing the alternatively spliced *MMSET* exon 4a[101,122]. The second transcription initiation site is within *MMSET* intron 9[188]. Translation of these transcripts initiates in *MMSET* exon 15 and produces the RE-IIBP protein, which is identical to the C-terminus of MMSET II.

The potential contribution of the MMSET protein variants is unclear. In 30-65% of patients, the breakpoints are downstream of the proper translation initiation site in *MMSET* exon 3[1,82,101,119]. Therefore, in patients with MB4-2 and MB4-3 breakpoints it is unknown if the overexpressed transcripts encoding MMSET I and MMSET II produce a protein product and if so how much of the N-terminus is lost. Alternative translation initiation sites in *MMSET* exons 4 and 6 were identified by Chesi et al.[91]. However, the N-terminal PWWP domain is truncated or lost entirely when either of these predicted sites are used. Moreover, if these alternative translation initiation sites are used, it is unclear if the truncated proteins produced are functionally equivalent to the wild-type variants overexpressed in MB4-1 patients.

The function of the various MMSET variants is currently unknown, however, based on the encoded protein domains the full length isoform, MMSET II, is likely a histone methyltransferase.[237] The overexpression of this isoform may significantly alter the expression of numerous genes by altering chromatin states across the genome or potentially in specific regions. The function of MMSET I is less clear. It may regulate the function of MMSET II by competing for target binding sites or essential co-factors. Limited information is available regarding the function of the RE-IIBP variant. It was originally identified in a library screen for proteins that bind the interleukin-5 response element II by Garlisi et al.[188]. Expression of RE-IIBP inhibited the expression of endogenous interleukin-5 and luciferase expression from a response element II reporter construct. Therefore, at the least, RE-IIBP is believed to be a negative regulator of interleukin-5 expression.

C3.2.1 - Working Hypothesis

The only universally overexpressed transcript from the MMSET locus that will encode a protein product with uncompromised function will be the transcript encoding RE-IIBP.

C3.3 – Chapter 3 Results

The majority of the results in this chapter are published in Blood in Keats et al. 2005[122]. However, the evolutionary analysis of MMSET is currently being prepared for submission as a separate manuscript.

C3.3.1 – MMSET Constructs

A series of MMSET expression constructs tagged with GFP at the N-terminus and C-terminus were constructed, which represent the universally dysregulated transcripts in t(4;14) positive patients with MB4-1, MB4-2, and MB4-3 breakpoints (Figure C3.1). The open reading frames encoding MMSET I, MMSET II, and RE-IIBP and their endogenous Kozak sequence was PCR amplified from plasmids containing the open reading frames of MMSET I or MMSET II provided by Leif Bergsagel (Mayo Clinic, Scottsdale). The respective open reading frames were cloned into the N-terminal and C-terminal GFP expression vectors of the Creator Cloning System, pLP-EGFP-C1 and pLPS-3'EGFP, respectively. To determine if the proposed alternative translation initiation sites in *MMSET* exons 4 and 6 are used, we amplified the open reading frame of MMSET I and MMSET II from the beginning of *MMSET* exons 4 and 5, respectively. These constructs, which represent the entire *MMSET* sequence overexpressed in MB4-2 and MB4-3 positive patients were cloned in to the C-terminal GFP expression vector pLPS-3'EGFP. A complete list of the cloned wild-type and predicted MB4-2 and MB4-3 protein variants are listed in Table C3.1 along with their respective size, isoelectic point, molecular weight, and molecular weight with GFP.

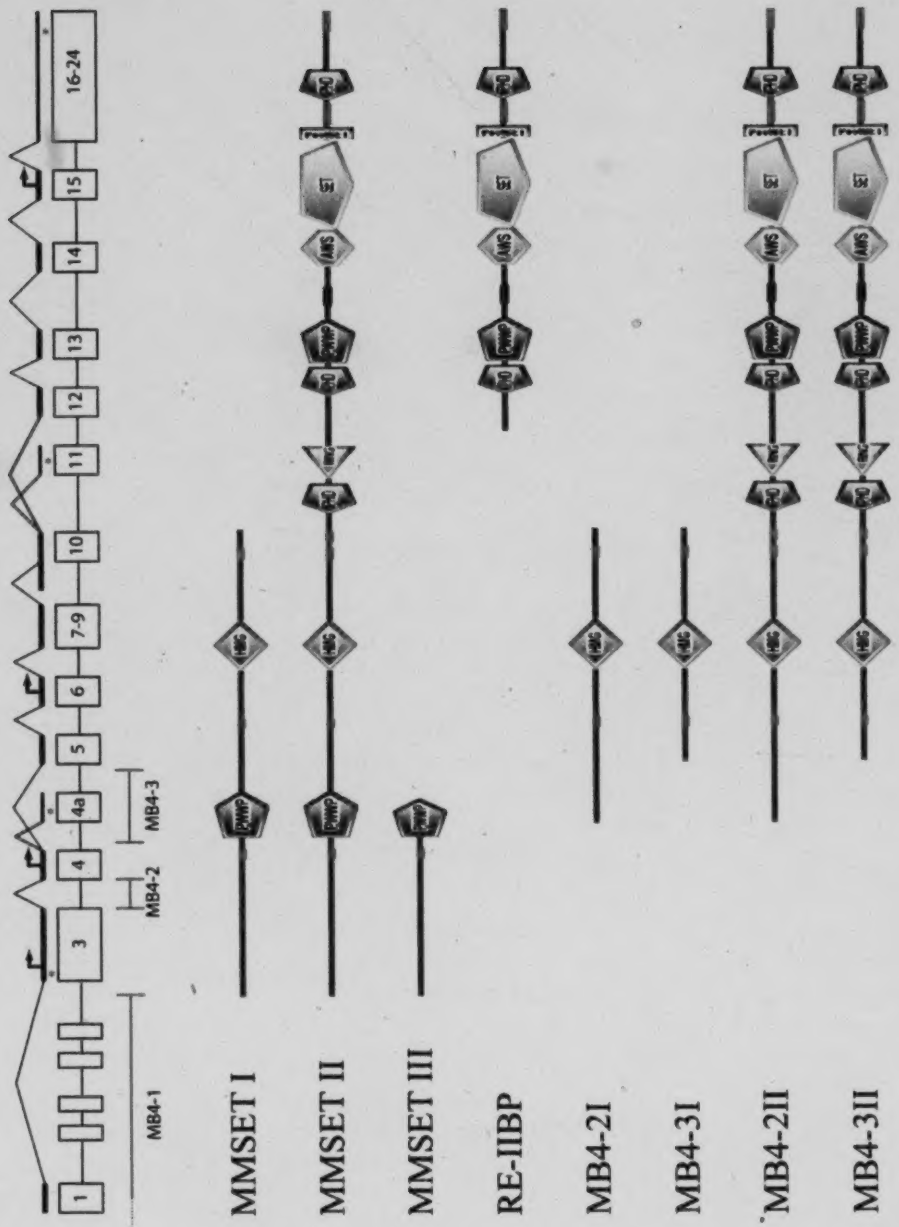

A
1
3
4
4a
5
6
7-9
10
11
12
13
14
15
16-24
MB4-1
MB4-2
MB4-3
B
MMSET I
MMSET II
MMSET III
RE-IIBP
MB4-2I
MB4-3I
MB4-2II
MB4-3II
PWWP
HMG
PHD
AWS
SET
PostSET

Figure C3.1 – MMSET Splicing and Protein Variants

A) An illustration with the exon-intron structure of *MMSET* is shown. The breakpoint cluster regions for MB4-1, MB4-2, and MB4-3 are shown. Thin black lines indicate the proper splicing pattern producing the mRNA species encoding MMSET II. Alternative splicing events which produced transcripts encoding MMSET III and MMSET I are indicated by thick red lines. In-frame stop codons are indicated by red asterisks. The proper *MMSET* translation initiation site is indicated by a green arrow in exon 3, while the alternative translation initiation sites in exon 4 and 6 identified by Chesi et al. are indicated by blue arrows[91]. The relative transcription initiation site for transcripts encoding RE-IIBP is indicated by a green line in intron 9 and the translation initiation site in exon 15 is indicated by a green arrow. B) The conserved protein domains present in wild-type and predicted MMSET protein variants identified by the S.M.A.R.T. protein prediction program[189,190]. For simplicity the predicted protein products in MB4-2 and MB4-3 positive patients are identified as MB4-2I, MB4-2II, MB4-3I, and MB4-3II to indicate truncated MMSET type I and II protein products where translation initiates in *MMSET* exon 4 or 6, respectively.

Table C3.1 – Features of MMSET Protein Variants

Protein Variant	Amino Acids	pI	MW (kDa)	MW with GFP
MMSET I	647	9.26	71.4	101.7
MMSET II	1365	9.00	152.3	182.6
MMSET III	273	9.38	30.2	60.5
RE-IIBP	584	8.92	66.4	96.7
MB4-2I	409	9.25	45.2	73.9
MB4-3I	324	9.05	35.6	64.3
MB4-2II	1127	8.95	126.1	154.8
MB4-3II	1042	8.89	116.4	145.1

The Mw and pI of each polypeptide was predicted with the compute pI/Mw tool on the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (http://au.expasy.org/tools/pi_tool.html).

C3.3.2 – Localization of Wild-type MMSET Protein Variants

The localization of MMSET I, MMSET II, and RE-IIBP was determined by transient transfection of HeLa cells with N-terminally tagged variants. This ensures the observed GFP localization is reflective of the entire wild-type protein, as translation initiates in GFP and extends to the end of the tagged protein. Moreover, the wild-type MMSET I and MMSET II proteins are predicted to be identical to the MMSET protein variants overexpressed in t(4;14) positive patients with MB4-1 breakpoints, because the translation initiation site in *MMSET* exon 3 is retained. As predicted, based on protein homology and the presence of nuclear localization signals, the wild-type/MB4-1 MMSET I and II variants localized to the nucleus, but they are excluded from nucleoli (Figure

C3.2). Unexpectedly, the localization of RE-IIBP was the exact inverse of the wild-type MMSET variants. RE-IIBP is enriched in two different compartments; in foci within the cytoplasm, and as a nuclear population localized almost exclusively to nucleoli (Figure C3.2).

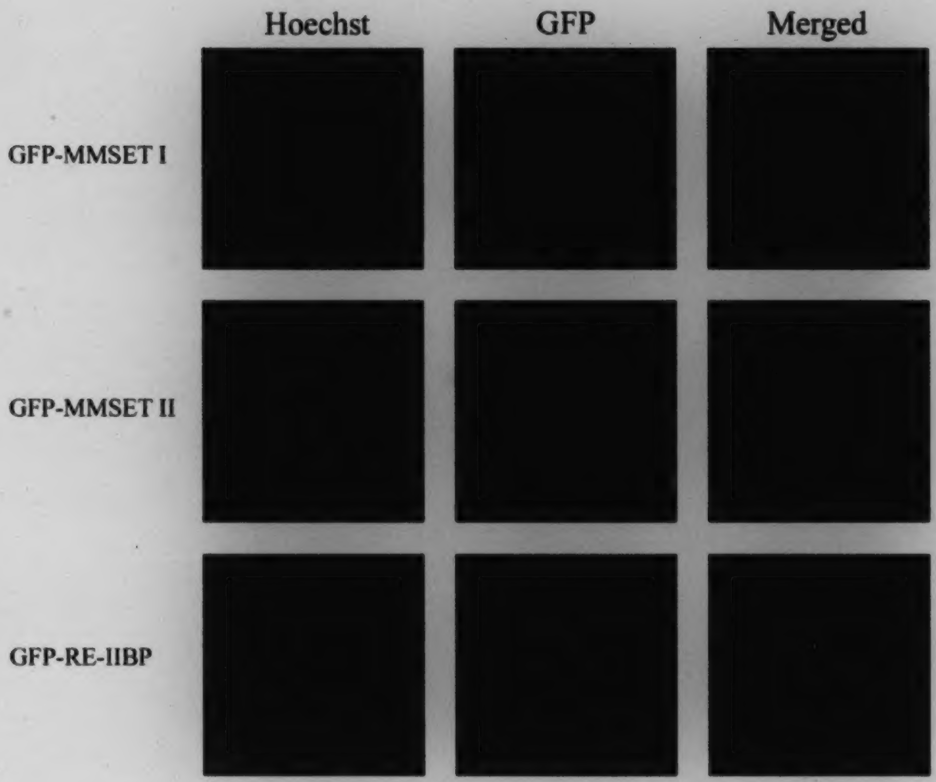


Figure C3.2 – Localization of N-Terminally Tagged Wild-type MMSET Variants

Live cell localization of MMSET I, MMSET II, and RE-IIBP tagged with GFP at the N-terminus in transiently transfected HeLa cells. The location of the nucleus and nucleoli are identified by the live cell permeable DNA dye, Hoechst.

To ensure the localization patterns seen with the N-terminal tags are not caused by a protein folding alteration caused by the tag, we additionally determined the localization of each variant with the alternative C-terminal tag. Some minor differences exist in the localization patterns between the N and C-terminally tagged variants (Figure C3.3). Interestingly, the localization patterns are more pronounced in constructs having the C-terminal tags. We believe this may reflect a lower level of expression due to differences in the translation efficiency of the MMSET and GFP Kozak sequences. The localization differences are most dramatic in the C-terminally tagged MMSET I and MMSET II variants. In C-terminally tagged variants the nucleoli contain a pale region of GFP fluorescence unlike the near perfect exclusion seen in the N-terminally tagged variants. In the case of RE-IIBP the localization pattern is identical, though the cytoplasmic foci are more pronounced. The RE-IIBP localization pattern is consistent for both N- and C-terminally tagged variants observed in live cells (Figure C3.2 and C3.3), in methanol or in paraformaldehyde fixed cells (data not shown).

Though localization does not necessarily reflect function, the localization pattern of MMSET II, which contains the histone methyltransferase SET domain, is almost identical to the staining of DNA/chromatin by Hoechst dye (Figure C3.3). Because an interaction with chromatin is likely a necessary aspect of proteins within this family, MMSET II may be a functional histone methyltransferase. Unlike MMSET II, the localization of MMSET I was often diffuse, though it may interact with DNA/chromatin albeit at a lower affinity.

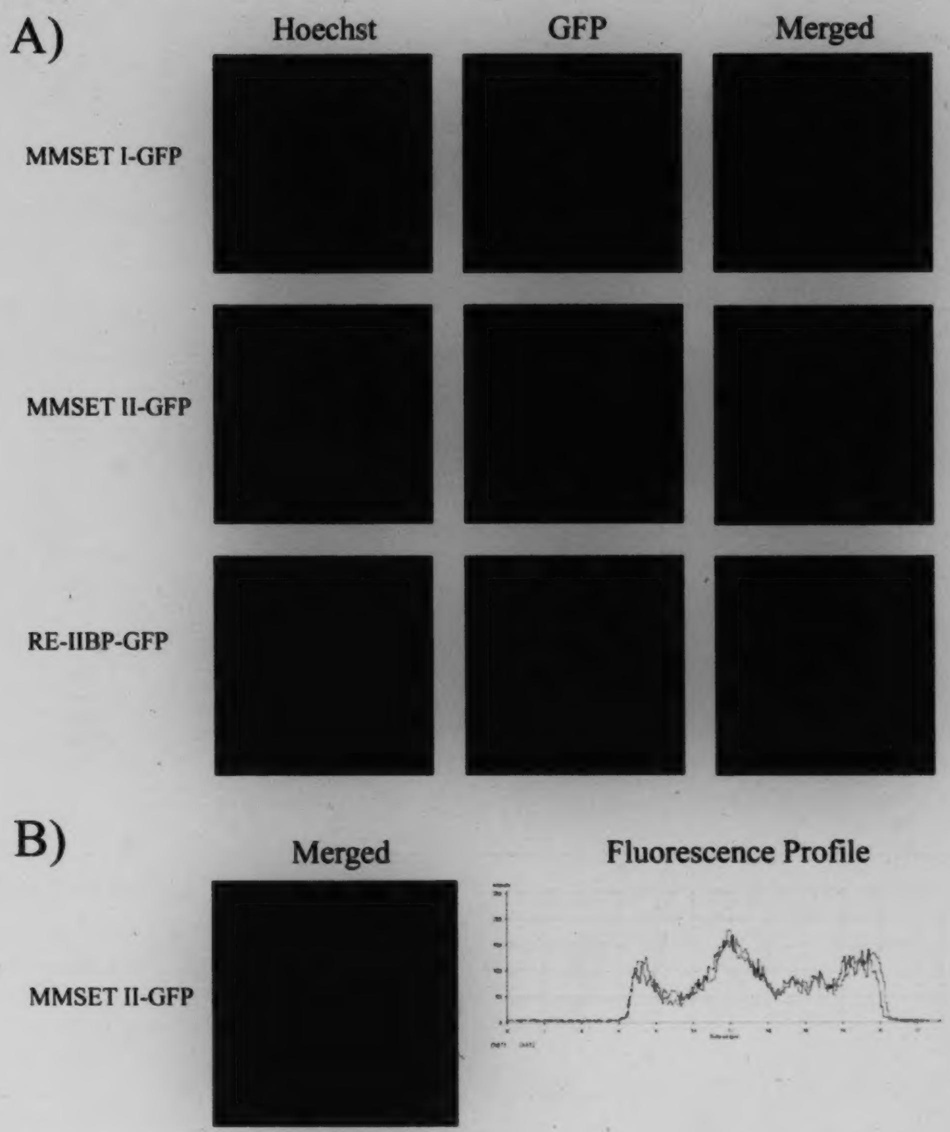

A)
Hoechst
GFP
Merged
MMSET I-GFP
MMSET II-GFP
RE-IIBP-GFP
B)
Merged
Fluorescence Profile
MMSET II-GFP

Figure C3.3 – Localization of C-terminally Tagged MMSET Protein Variants

A) Live cell localization of MMSET I, MMSET II, and RE-IIBP tagged with GFP at the C-terminus. B) Co-localization of MMSET II-GFP with DNA/chromatin. The fluorescence profile is generated from the area covered by the red arrow. The blue plot represents the intensity of the Hoechst stain and the green plot represents the intensity of MMSET II-GFP.

C3.3.3 – Localization of Myeloma Breakpoint Specific MMSET Protein Variants

In t(4;14) positive myeloma patients with MB4-2 and MB4-3 breakpoints, translation of the hybrid transcripts and/or *de novo* transcripts from secondary translation initiation sites are predicted to produce truncated proteins lacking the N-terminus of MMSET, as the wild-type translation initiation site is lost (Figure C3.1). To determine if hybrid transcripts from each breakpoint variant can produce protein products, we cloned the type I and II breakpoint variants from the beginning of *MMSET* exon 4 and 5 to mimic the hybrid transcripts present in MB4-2 and MB4-3 patients, respectively. These cloned fragments, lacking the proper *MMSET* translation initiation site in exon 3, were cloned into a C-terminal GFP tag vector and anti-GFP immunoblots were performed on transiently transfected HeLa cells to identify protein products originating from secondary translation initiation sites. We detected truncated protein products reflecting the predicted protein size of the various polypeptides produced from the alternative translation initiation sites in *MMSET* exons 4 and 6 in HeLa cells transiently transfected with type I and II breakpoint variant constructs (Figure C3.4). Interestingly, two bands corresponding to the predicted proteins products of the alternative translation initiation sites in *MMSET* exons 4 and 6 are detected in the MB4-2I construct, suggesting the alternative translation initiation site in *MMSET* exon 4 is not preferentially used when this is the first available site. Identical results are seen in myeloma cell lines with MB4-2 and MB4-3 breakpoints using an antibody raised against the C-terminus of MMSET II (Leif Bergsagel, Pers. Comm.). Furthermore, the translation initiation site in *MMSET* exon 15 used to produce RE-IIBP is functional based on the fact that C-terminal tagged constructs of RE-IIBP were detected. This confirms the detection of endogenous RE-IIBP in human T-cell nuclear extracts by Garlisi et al.[188]. Therefore, the RE-IIBP localization pattern observed with the GFP constructs is likely representative of the endogenous protein.

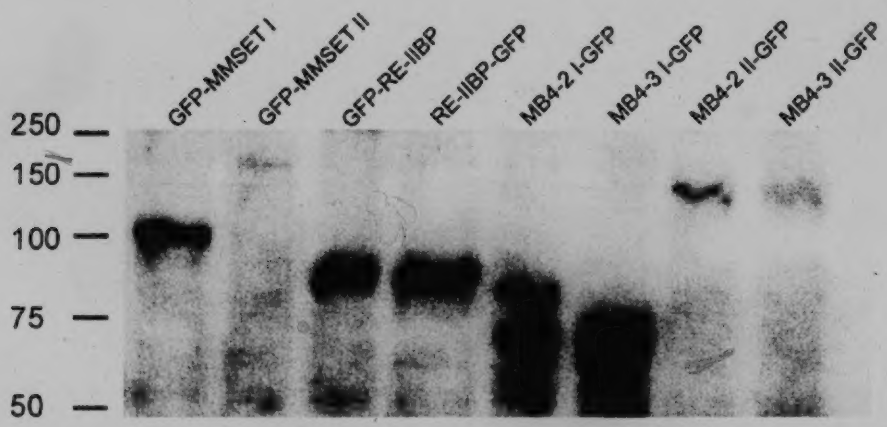


Figure C3.4 – Truncated Protein Products are Produced from the Myeloma Specific MB4-2 and MB4-3 Transcripts.

An anti-GFP immunoblot of isolated proteins from HeLa cells transiently transfected with various MMSET constructs is shown. GFP-protein indicates N-terminally tagged constructs and protein-GFP indicates C-terminally tagged constructs. The detection of the predicted protein products in the C-terminally tagged MB4-2, MB4-3, and RE-IIBP constructs confirms the alternative translation initiation sites in exon 4, 6, and 15 are functional. Interestingly, the MB4-2 construct resulted in two protein products, indicating that neither alternative translation site is dominant.

The localization of the truncated MB4-2 and MB4-3 breakpoint variants were determined in transiently transfected HeLa cells using the C-terminal tagged constructs.

All four variants, both type I and II from each of the MB4-2 and MB4-3 genomic breakpoints, localized to the nucleus (Figure C3.5). However, unlike the wild-type/MB4-1 proteins with the N-terminus of MMSET, all four breakpoint variants are enriched in nucleoli (Figure C3.5). Interestingly, the only conserved protein domain which is lost in the truncated MB4-2 and MB4-3 variants is the N-terminal PWWP domain (Figure C3.1). In MB4-3 variants the entire domain is lost, while in MB4-2 variants the C-terminal region is retained but the P-W-W-P motif is lost. Therefore, either the N-terminal PWWP domain or an as yet unidentified domain in the N-terminus of MMSET regulates its localization. Importantly, the localization differences between the wild-type/MB4-1 variants and the truncated variants produced from MB4-2 and MB4-3 transcripts suggests that the different breakpoint variants are unlikely to be functionally equivalent.

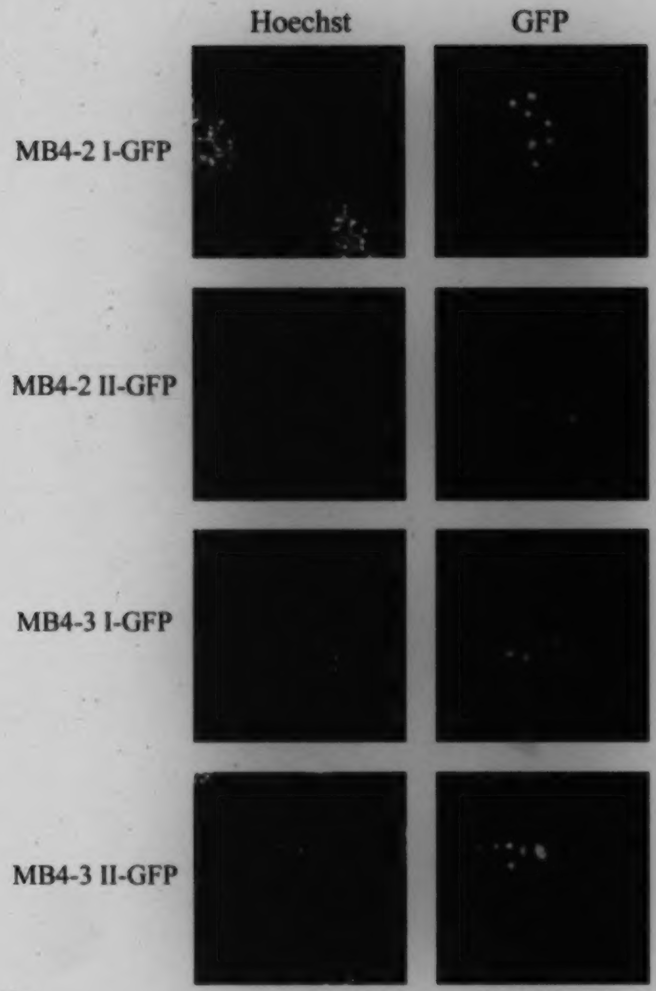

Hoechst
GFP
MB4-2 I-GFP
MB4-2 II-GFP
MB4-3 I-GFP
MB4-3 II-GFP

Figure C3.5 – Alternative Translation Sites Produce Mis-Localized MMSET Variants

Live cell localization of C-terminally tagged type I and II constructs, corresponding to the MMSET transcripts detected in MB4-2 and MB4-3 patients, in transiently transfected HeLa cells. The live cell permeable DNA stain, Hoechst, is used to identify the nucleus and nucleoli. All four novel MMSET constructs, which are unique to the MB4-2 and

MB4-3 breakpoint variants, result in protein products that enrich in nucleoli unlike the wild-type variants that are excluded from nucleoli.

C3.3.4 – Characterization of a Novel Localization Domain within MMSET III

The different localization patterns identified between the wild-type/MB4-1 and breakpoint specific variants appears to be related to the loss of the N-terminus of MMSET. This N-terminal region is encoded by *MMSET* exons 3 and 4 which are partially or completely lost in the MB4-2 and MB4-3 protein variants, respectively. To determine if this small N-terminal region regulates the nucleolar exclusion of wild-type/MB4-1 MMSET variants, we cloned the MMSET III open reading frame, which contains the alternatively spliced *MMSET* exon 4a identified by Malgeri et al. from one of our t(4;14) positive patients with an MB4-1 breakpoint (NCBI Accession Number AY694128)(Figure C3.1)[101]. This novel splice variant encodes a truncated protein largely representing the N-terminal coding region deleted in the MB4-2 breakpoint variants since MMSET exon 4a contains an in-frame stop codon (Figure C3.1). The predicted polypeptide shares 15 amino acids with the truncated MB4-2 protein variants and though a portion of the N-terminal PWWP domain is lost, the P-W-W-P sequence motif is maintained in MMSET III. The localization of MMSET III in transiently transfected HeLa cells was similar to the wild-type/MB4-1 protein variants, nuclear and excluded from nucleoli, regardless of which terminus was tagged with GFP (Figure C3.6). Interestingly, this variant does not contain a previously identified nuclear localization signal[90,91], nor have we identified one using online prediction programs.

To determine if the N-terminus of MMSET characterized by MMSET III could alter the localization of a nucleolar protein, we cloned MMSET III into a GFP-B23 expression vector provided by Michael Hendzel (University of Alberta/Cross Cancer Institute, Edmonton). B23 is a nucleolar protein involved in ribosome biogenesis[238]. Transient transfection of HeLa cells with the GFP-MMSET III-B23 construct resulted in a mixed population of GFP positive cells (Figure C3.6). A blinded quantitation of 500 cells from random fields of view showed 53.1% had a wild-type MMSET phenotype (excluded from nucleoli), 41.5% a B23 phenotype (nucleolar), and 5.4% an unclassifiable phenotype (diffuse nuclear staining in both compartments). Therefore, the N-terminus of MMSET characterized by MMSET III contains a domain regulating its exclusion from nucleoli and this domain can even prevent a nucleolar protein from localizing to the nucleolus. However, it is not known if an as yet unidentified N-terminal domain or the partially retained N-terminal PWWP domain regulates this localization.

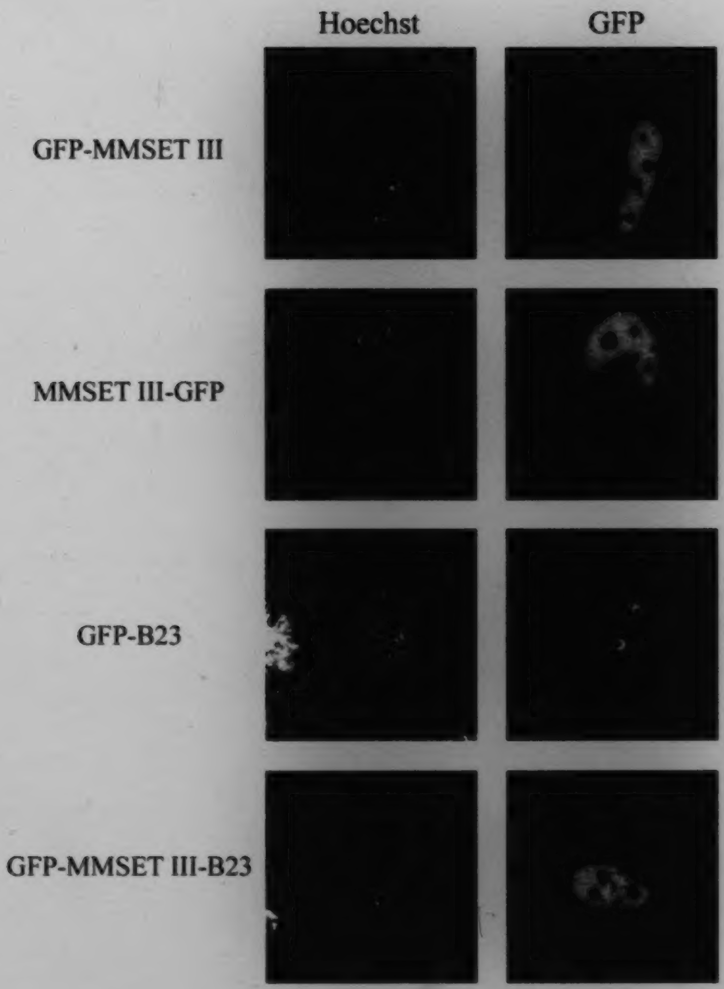


Figure C3.6 – The N-terminus of MMSET regulates its Localization Pattern.

Live cell localization of N- and C-terminally GFP tagged MMSET III, and N-terminally GFP tagged B23 and MMSET III-B23 hybrid constructs in transiently transfected HeLa cells. The live cell permeable DNA stain, Hoechst, is used to identify the nucleus and nucleoli. The MMSET III constructs shows a nuclear localization with a nucleolar exclusion pattern. The B23 construct shows the characteristic nucleolar localization pattern while the MMSET III-B23 hybrid construct shows the typical MMSET phenotype (nuclear and excluded from nucleoli). Immunoblotting experiments of the MMSET III-B23 constructs confirmed that only the hybrid GFP-Exon 4a-B23 protein product is produced by this expression vector (Not Shown).

C3.3.5 – Functional Assessment of MMSET Protein Variants

The localization differences between the wildtype/MB4-1 and MB4-2/MB4-3 breakpoint variants suggested a loss of function for MB4-2 and MB4-3 specific protein variants. Since the function of MMSET is unknown, we indirectly tested this hypothesis using a technique called fluorescence recovery after photobleaching (FRAP). FRAP involves the bleaching of a small region of fluorescence and recording over time the recovery of fluorescence into the bleached region, as bleached molecules are replaced by unbleached molecules, to determine the time to half recovery ($t_{1/2}$). FRAP recovery kinetics are mediated by size, cellular compartment, compartment viscosity, affinity of protein-protein interactions (binding/unbinding events), protein-protein/structure collisions, and temperature[239,240].

The FRAP recovery kinetics of the N-terminal tagged MMSET variants showed the predicted results. As the protein size increased; MMSET III, MMSET I, and MMSET II, the $t_{1/2}$ times increased, due in part, to diffusion kinetics of larger molecules (Table C3.2). However, the $t_{1/2}$ of MMSET II, compared to MMSET I, was much slower (150 vs 4 seconds) than predicted by simple diffusion kinetics. The differential FRAP kinetics suggest MMSET II but not MMSET I is binding to a nucleoplasmic structure with very high affinity. The predicted function of MMSET II as a histone methyltransferase and the co-localization with DNA/chromatin suggests that this nucleoplasmic structure is most likely chromatin.

Table C3.2 – FRAP $t_{1/2}$ Times of the Various MMSET Variants

	N-terminal GFP $t_{1/2}$ Time (Sec)	C-terminal GFP $t_{1/2}$ Time (Sec)
MMSET I	$4.0^{+/-}0.07$	$2.957^{+/-}0.06$
MMSET II	$150.0^{+/-}0.08$	$130.0^{+/-}0.06$
MMSET III	$0.986^{+/-}0.12$	-----
MB4-2 I	-----	$0.772^{+/-}0.05$
MB4-2 II	-----	$19.0^{+/-}0.07$
MB4-3 I	-----	$0.772^{+/-}0.03$
MB4-3 II	-----	$12.0^{+/-}0.05$

The mean $t_{1/2}$ recovery time for each construct is shown plus/minus the standard deviation. The bleached ROI, for each FRAP experiment from the different MMSET variants, was placed specifically in the nucleoplasm to measure the kinetic recovery of the nucleoplasmic pool of each variant.

The C-terminal tagged MB4-2 and MB4-3 breakpoint variants were compared to the similarly tagged wildtype/MB4-1 MMSET variants. For both MMSET I and II constructs, the $t_{1/2}$ was slightly faster for the C-terminal tags as compared to the N-terminal tags, however, the difference was not statistically significant (Table C3.2). The recovery of the truncated MB4-2 and MB4-3 variants was substantially faster than their wild-type/MB4-1 counterparts ($P<0.001$) (Table C3.2)(Figure C3.7). The N-terminus of MMSET drastically affects the mobility of MMSET variants, based on the observation that the $t_{1/2}$ of the type I constructs decreased to 26% and the type II constructs decreased to 14.6% and 9.2% of the wild-type/MB4-1 protein. Moreover, the N-terminus appears to be a vital part of a synergistic interaction involving multiple component domains of MMSET, as even the additive $t_{1/2}$ of MMSET I and MB4-2 II, which represents the entire

MMSET II protein, does not match the $t_{1/2}$ of MMSET II (21.96 versus 130.00 seconds). Therefore the N-terminus of MMSET characterized by MMSET III is essential for the localization and function of MMSET I and MMSET II.

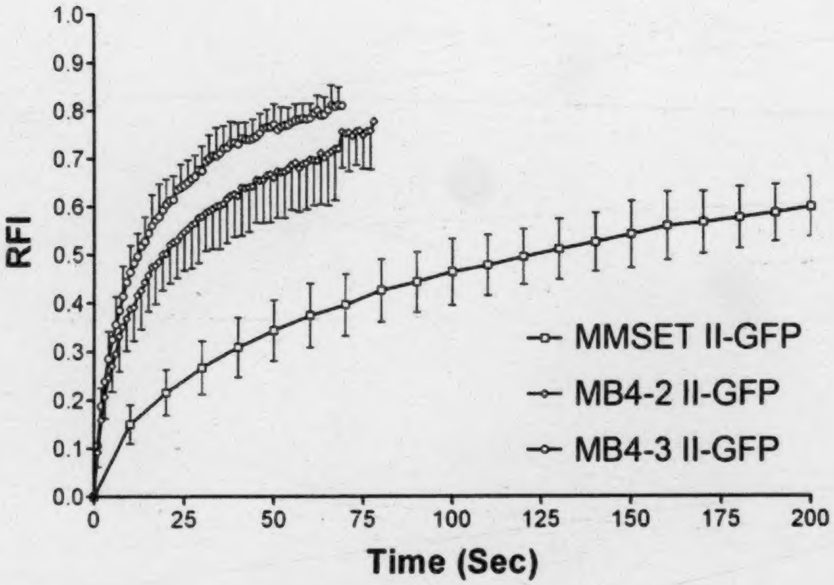
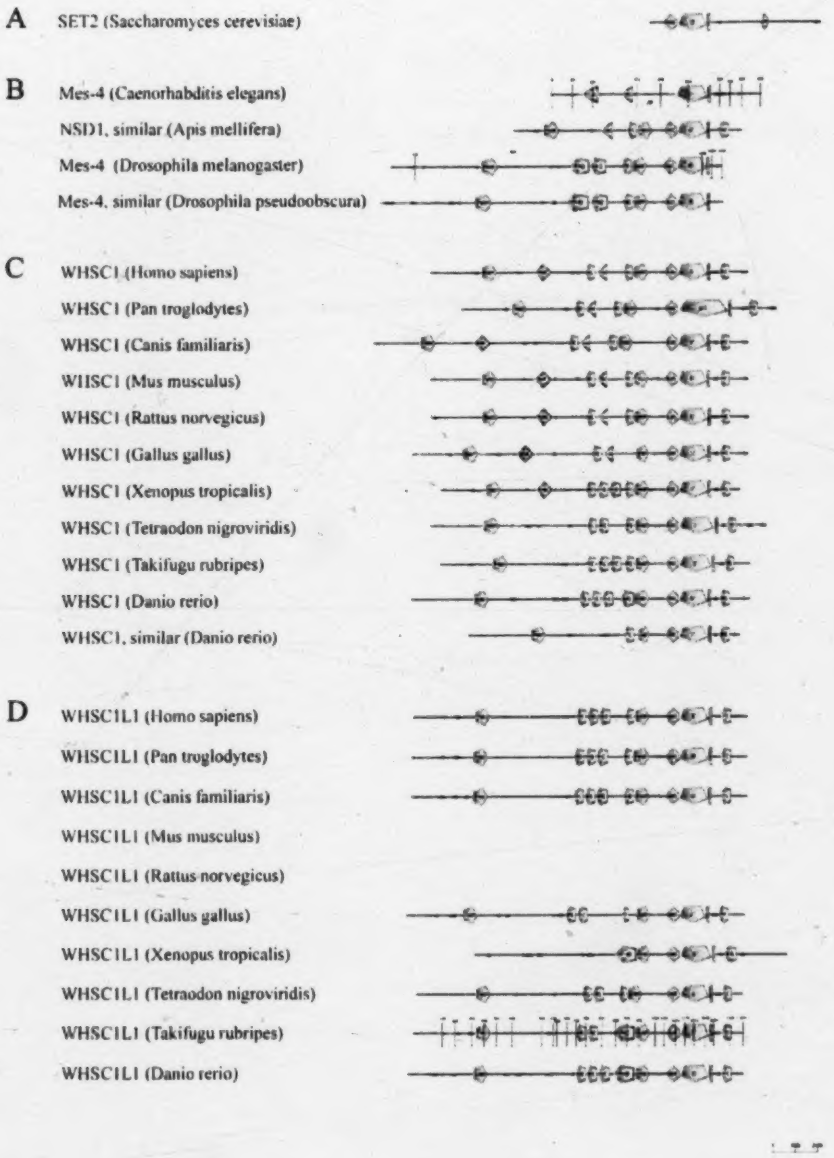


Figure C3.7 – FRAP Recovery Curves of Wild-type and Breakpoint Variants. FRAP recovery curves from nucleoplasmic bleaching experiments of C-terminally tagged MMSET II/MB4-1 II, MB4-2 II, and MB4-3 II protein variants. Error bars represent the standard deviation. Only the initial recovery of MMSET II is represented as the x-axis has been shortened to allow the differences in the $t_{1/2}$ times to be more evident.

C3.3.6 – Evolutionary Analysis of MMSET

The localization and FRAP kinetics of the various MMSET protein variants suggests that an essential domain exists within the N-terminus of wild-type MMSET protein variants. Since this region likely encodes a domain essential for the localization and function of MMSET we hypothesized that the region would be conserved in evolution. Database searches within the NCBI Homologene Database (http://www.ncbi.nlm.nih.gov), the Ensembl Genome Browser (http://www.ensembl.org), and BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/) with the N-terminus and C-terminus of MMSET identified a large number of NSD paralogs and orthologs within 15 different species from yeast to humans. The predicted protein homologies were verified by three independent methods. First, the presence of the predicted *FGFR* and/or *TACC* paralog on the same genomic contig was confirmed. Second, the protein sequence was analyzed with the S.M.A.R.T. protein prediction program to verify the predicted protein domain architecture of each paralog. Third, all of the protein sequences were aligned and a nearest-neighbor phylogeny test was conducted to verify the groupings. These criteria were used to separate the identified proteins into paralog families (*WHSC1/MMSET/NSD2*, *WHSC1L1/NSD3*, and *NSD1*) conserved through evolution (Figure C3.8). Based on this grouping there is only a single NSD ortholog in *Saccharomyces cerevisiae* (SET2), *Caenorhabditis elegans* (Mes-4), or the taxonomy class insecta (Figure C3.9). However, during the evolution of chordates the single NSD ortholog was duplicated twice to generate the human *WHSC1/MMSET/NSD2*, *WHSC1L1/NSD3*, and *NSD1* gene orthologs (Figure C3.9). Interestingly, within the taxonomy class actinopterygii an additional duplication of the *NSD1* ortholog occurred as both pufferfish and zebrafish have two *NSD1* similar genes. Moreover, in zebrafish (*Danio rerio)* a second duplication of the *WHSC1/MMSET/NSD2* ortholog occurred resulting in two *WHSC1* similar genes.


A
SET2 (Saccharomyces cerevisiae)
B
Mes-4 (Caenorhabditis elegans)
NSD1, similar (Apis mellifera)
Mes-4 (Drosophila melanogaster)
Mes-4, similar (Drosophila pseudoobscura)
C
WHSC1 (Homo sapiens)
WHSC1 (Pan troglodytes)
WHSC1 (Canis familiaris)
WHSC1 (Mus musculus)
WHSC1 (Rattus norvegicus)
WHSC1 (Gallus gallus)
WHSC1 (Xenopus tropicalis)
WHSC1 (Tetraodon nigroviridis)
WHSC1 (Takifugu rubripes)
WHSC1 (Danio rerio)
WHSC1, similar (Danio rerio)
D
WHSC1L1 (Homo sapiens)
WHSC1L1 (Pan troglodytes)
WHSC1L1 (Canis familiaris)
WHSC1L1 (Mus musculus)
WHSC1L1 (Rattus norvegicus)
WHSC1L1 (Gallus gallus)
WHSC1L1 (Xenopus tropicalis)
WHSC1L1 (Tetraodon nigroviridis)
WHSC1L1 (Takifugu rubripes)
WHSC1L1 (Danio rerio)

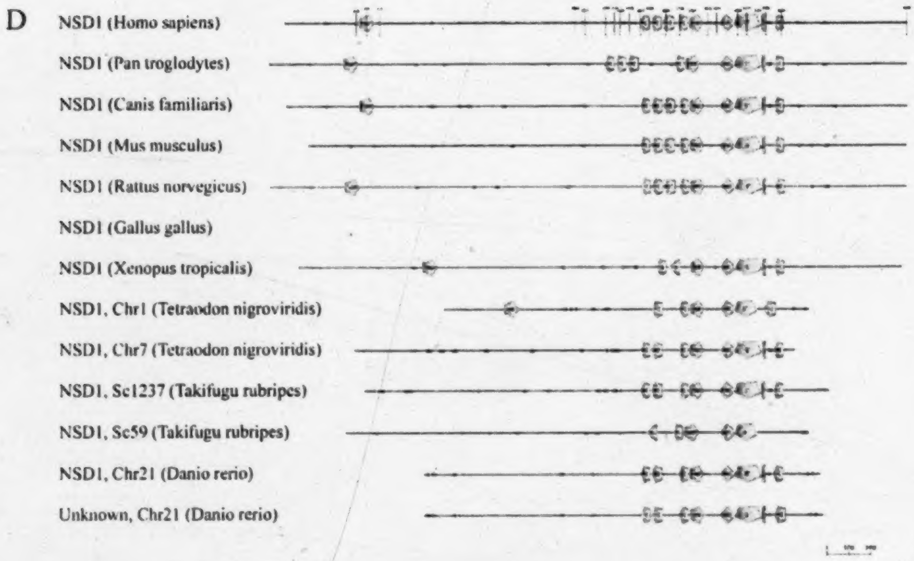

D
NSD1 (Homo sapiens)
NSD1 (Pan troglodytes)
NSD1 (Canis familiaris)
NSD1 (Mus musculus)
NSD1 (Rattus norvegicus)
NSD1 (Gallus gallus)
NSD1 (Xenopus tropicalis)
NSD1, Chr1 (Tetraodon nigroviridis)
NSD1, Chr7 (Tetraodon nigroviridis)
NSD1, Sc1237 (Takifugu rubripes)
NSD1, Sc59 (Takifugu rubripes)
NSD1, Chr21 (Danio rerio)
Unknown, Chr21 (Danio rerio)

Figure C3.8 – Domain Architecture of NSD Paralogs and Orthologs

The scale protein predictions generated by the SMART protein prediction program are shown. A) The single NSD1 ortholog identified in Saccharomyces cerevisiae. B) The single NSD ortholog identified in non-chordates. C) The *WHSC1/MMSET/NSD2* orthologs identified in chordates. D) The *WHSC1L1/NSD3* orthologs identified in chordates. E) The *NSD1* orthologs identified in chordates. The gene identification tags and protein sequences retrieved from the NCBI and Ensembl databases for this analysis are as follows: *Saccharomyces cerevisiae* SET1 (NP_011987.1); *Apis mellifera* NSD1 similar/LOC412225 (XP_395687); *Caenorhabditis elegans* Mes-4 (NP_506333); *Drosophila melanogaster* Mes-4/CG4976-PA (NP_733239.1), *Drosophila pseudoobscura* Mes-4 similar/GA18567-PA (EAL27392); *Homo sapiens WHSC1* (NP_579890.1), *WHSC1L1* (NP_075447.1), *NSD1* (NP_071900.2); *Pan troglodytes* WHSC1/LOC461070 (XP_517068), WHSC1L1/ENSPTRG00000020173

(ENSPTRP00000034535), NSD1/LOC471754 (XP_527132); *Canis familiaris* WHSC1/LOC479077 (XP_536224.1), WHSC1L1/LOC475584 (XP_532803), NSD1/ENSCAFG00000016473 (ENSCAFP00000024244); *Mus musculus Whsc1* (XP_132006.4), *Nsd1* (NP_032765); *Rattus norvegicus* WHSC1/LOC305456 (XP_223540), NSD1/LOC306764 (XP_225168); *Gallus gallus* WHSC1/LOC422897 (XP_420839), WHSC1L1/LOC426778 (XP_424390); *Xenopus tropicalis* WHSC1/ENSXETG00000002393 (ENSXETP00000005100), WHSC1L1 (GENSCAN00000113291), NSD1/ENSXETG00000007233 (GENSCAN00000091765); *Tetraodon nigroviridis* WHSC1/GSTENG00023330001 (GSTENP00023330001), WHSC1L1/GSTENG00029589001 (GSTENP00029589001), NSD1, chr 1 (GSCT00001602001), NSD1, chr 7/GSTENG00014627001 (GIDT00013147001); *Takifugu rubripes* WHSC1/SINFRUG00000138891 (GENSCAN00000001541), WHSC1L1/SINFRUG00000145857 (SINFRUP00000154949), NSD1 Sc1237/SINFRUG00000135649 (GENSCAN00000017518), NSD1 Sc59/SINFRUG00000143574 (GENSCAN00000033944), *Danio rerio* WHSC1 (GENSCAN00000001643), WHSC1 similar (GENSCAN00000011251), WHSC1L1/ENSDARG00000007772 (ENSDARP00000044502), NSD1 Chr 21/ENSDARG00000030875 (ENSDARP00000036657), Unknown, Chr 21 (GENSCAN00000016175).

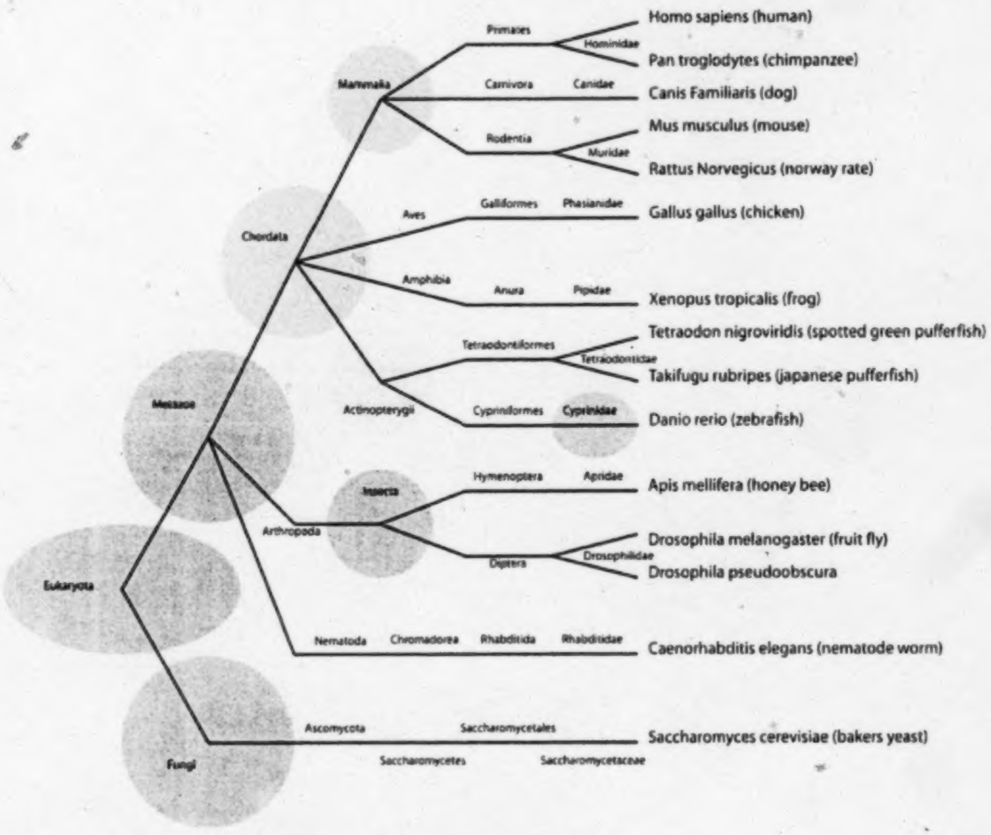

Homo sapiens (human)
Pan troglodytes (chimpanzee)
Canis Familiaris (dog)
Mus musculus (mouse)
Rattus Norvegicus (norway rate)
Gallus gallus (chicken)
Xenopus tropicalis (frog)
Tetraodon nigroviridis (spotted green pufferfish)
Takifugu rubripes (japanese pufferfish)
Danio rerio (zebrafish)
Apis mellifera (honey bee)
Drosophila melanogaster (fruit fly)
Drosophila pseudoobscura
Caenorhabditis elegans (nematode worm)
Saccharomyces cerevisiae (bakers yeast)
Primates
Hominidae
Mammalia
Carnivora
Canidae
Rodentia
Muridae
Aves
Galliformes
Phasianidae
Chordata
Amphibia
Anura
Pipidae
Tetraodontiformes
Tetraodontidae
Metazoa
Actinopterygii
Cypriniformes
Cyprinidae
Insecta
Hymenoptera
Apidae
Arthropoda
Diptera
Drosophilidae
Eukaryota
Nematoda
Chromadorea
Rhabditida
Rhabditidae
Ascomycota
Saccharomycetales
Saccharomycetes
Saccharomycetaceae
Fungi

Figure C3.9 – The Evolution of the NSD Family

An evolutionary tree illustrating the taxonomy of each organism with identified NSD family members is shown. Organisms contained within the grey circles have a single NSD ortholog. Organisms within the phylum chordata, identified by a pink circle, have at least three NSD paralogs representing the gene orthologs of *WHSC1/MMSET/NSD2*, *WHSC1L1/NSD3*, and *NSD1*. Organisms within the class actinopterygii, identified by a green circle, have at least four NSD paralogs resulting from a duplication of the *NSD1* ortholog. Moreover, in *Danio rerio* (zebrafish) an additional duplication of the *WHSC1* ortholog, indicated by a purple circle, results in five different NSD paralogs.

Previous work by Kouzarides identified four principle histone methyltransferase families in humans; SET1, SET2, SUV39, and RIZ defined by the nearest ortholog in *Saccharomyces cerevisiae*[193]. The SET2 family includes *WHSC1/MMSET/NSD2*, *WHSC1L1/NDS3*, *NSD1*, *ASH1L*, and *HYPB*. Through our database searches we identified a single NSD ortholog in *Drosophila melanogaster* (Mes-4/CG4976-PA), along with the previously identified *ASH1L* ortholog Ash1/CG8887-PA[241]. Since the NSD ortholog is duplicated twice in humans we searched the available human genome resources for additional *ASH1L* and *HYPB* paralogs, however, none were identified. Moreover, to rule out *HYPB* and *ASH1L* being paralogs derived from the ash1/CG4976-PA gene we searched the *Drosophila melanogaster* genome for a *HYPB* ortholog and found HYPB, like/CG1716-PA. Therefore, the SET2 gene appears to have diverged into three genes in *Drosophila melanogaster* (Mes-4, Ash1, and HYPB, like) with the subsequent divergence of Mes-4 into three additional SET2 family members in humans (Figure C3.10).

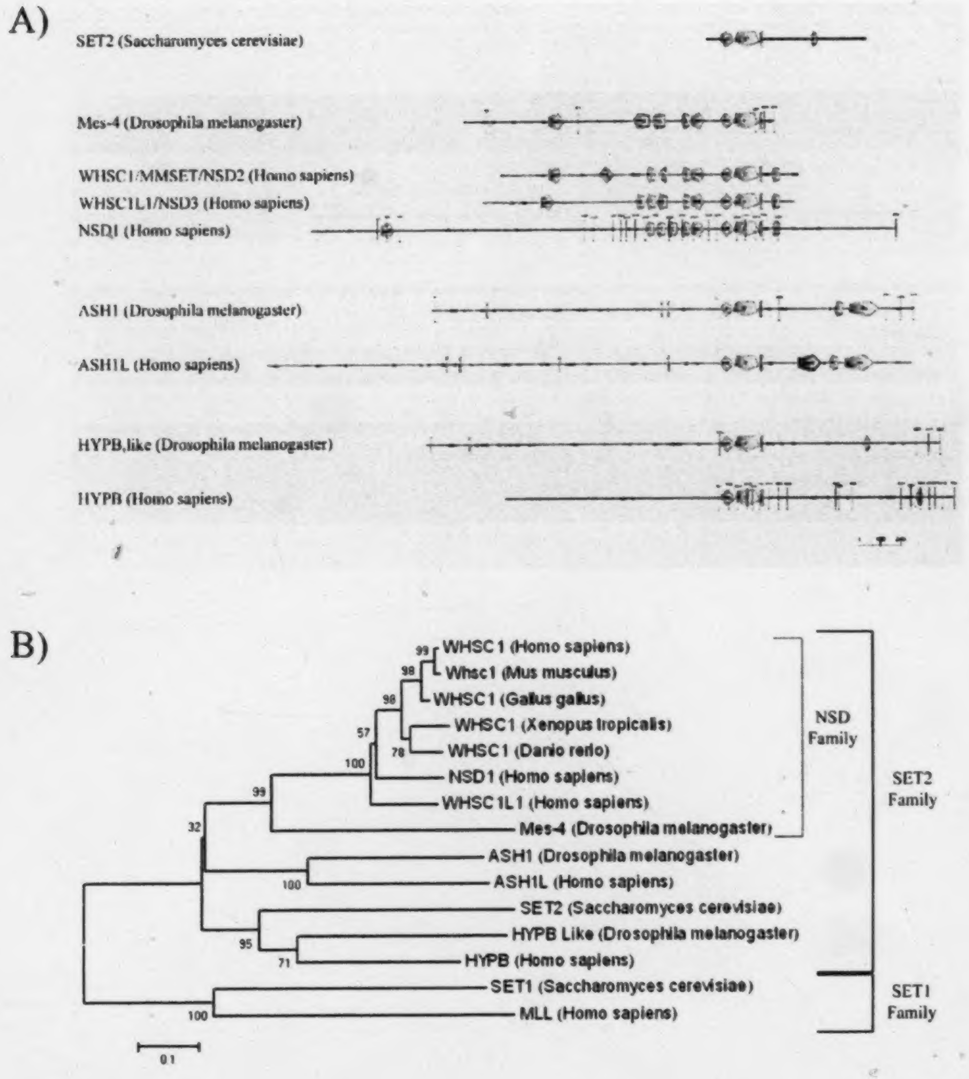

A)
SET2 (Saccharomyces cerevisiae)
Mes-4 (Drosophila melanogaster)
WHSC1/MMSET/NSD2 (Homo sapiens)
WHSC1L1/NSD3 (Homo sapiens)
NSD1 (Homo sapiens)
ASH1 (Drosophila melanogaster)
ASH1L (Homo sapiens)
HYPB,like (Drosophila melanogaster)
HYPB (Homo sapiens)
B)
99
WHSC1 (Homo sapiens)
98
Whsc1 (Mus musculus)
98
WHSC1 (Gallus gallus)
57
WHSC1 (Xenopus tropicalis)
78
WHSC1 (Danio rerio)
100
NSD1 (Homo sapiens)
99
WHSC1L1 (Homo sapiens)
32
Mes-4 (Drosophila melanogaster)
ASH1 (Drosophila melanogaster)
100
ASH1L (Homo sapiens)
SET2 (Saccharomyces cerevisiae)
95
HYPB Like (Drosophila melanogaster)
71
HYPB (Homo sapiens)
SET1 (Saccharomyces cerevisiae)
100
MLL (Homo sapiens)
0.1
NSD Family
SET2 Family
SET1 Family

Figure C3.10 – Evolution of the SET2 Histone Methyltransferase Family

A) Scale protein predictions generated by the SMART protein prediction program for the various SET2 histone methyltransferase family members in *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Homo sapiens* are shown. During the evolution from unicellular to multi-cellular organisms the SET2 gene was duplicated twice to generate three related genes in *Drosophila melanogaster*. Subsequently, in humans the Mes-4 gene was duplicated two more times to generate *WHSC1/MMSET/NSD2*, *WHSC1L1/NSD3*, and *NSD1*. B) A multiple sequence alignment of the protein sequences encoding the AWS, SET, and post SET domains identified by the SMART protein prediction program was generated with ClustalW (http://www.ebi.ac.uk/clustalw/) using default parameters. A phylogeny tree was constructed from the multiple sequence alignment with the MEGA 2.1 free-ware package (www.megasoftware.net). The tree was constructed with the neighbor-joining method using the poisson correction model, handling missing data/gaps by pair-wise deletion, and a 500 replication bootstrap test of phylogeny. The SET1, SET2, and NSD histone methyl-transferase gene families are indicated. The protein sequences used for this analysis are as follows: *Saccharomyces cerevisiae* SET1 (NP_011987.1), SET2 (NP_012367.1); *Drosophila melanogaster* Mes-4/CG4976-PA (NP_733239.1), ASH1/CG8887-PA (NP_524160.1), HYPB, like/CG1716-PA (NP_572888); *Homo sapiens HYPB* (NP_054878.3), *ASH1L* (NP_060959.1), *WHSC1* (NP_579890.1), *WHSC1L1* (NP_075447.1), *NSD1* (NP_071900.2); *Mus musculus Whsc1* (XP_132006.4); *Xenopus tropicalis* WHSC1/ENSXETG00000002393 (ENSXETP00000005100); *Gallus gallus* WHSC1/LOC422897 (XP_420839); *Danio rerio* WHSC1 (GENSCAN00000001643).

The only conserved protein domain in the N-terminus of wild-type MMSET proteins which is lost in the breakpoint variants is the PWWP domain. However, based on the sequence alignment of MMSET I, MMSET II, and MB4-2 I the N-terminal PWWP domain is unlikely to be the domain regulating the localization of wild-type

MMSET proteins (Figure C3.11). The PWWP domain is a conserved protein domain of 50-80 amino acids[192]. However, the exact C-terminal boundary of the PWWP domain is not well defined. The N-terminal boundary is well defined as several proteins have PWWP domains at their N-termini. Unfortunately, the C-terminus is less well defined and at least for the PWWP domain in the *Saccharomyces pombe* gene SPBC215.07c an additional 30 C-terminal amino acids are required for proper folding[208]. Therefore, at a minimum the entire defined PWWP domain is likely essential for the proper function of the domain. As such, the loss of the highly conserved P-W-W-P motif in the MB4-2 protein variants and approximately half of the C-terminus of the domain in MMSET III, suggests the PWWP domain is not functional in either of these variants (Figure C3.11). Therefore, the domain regulating the localization of wild-type MMSET variants is likely N-terminal of the PWWP domain. To identify this domain we first aligned the N-termini of WHSC1/MMSET and the most homologous paralog WHSC1L1. This alignment identified two regions of high identity with one being the PWWP domain as expected (Figure C3.11). The other region of high identity encompasses amino acids 116-146 of MMSET I or MMSET II (Figure C3.11).

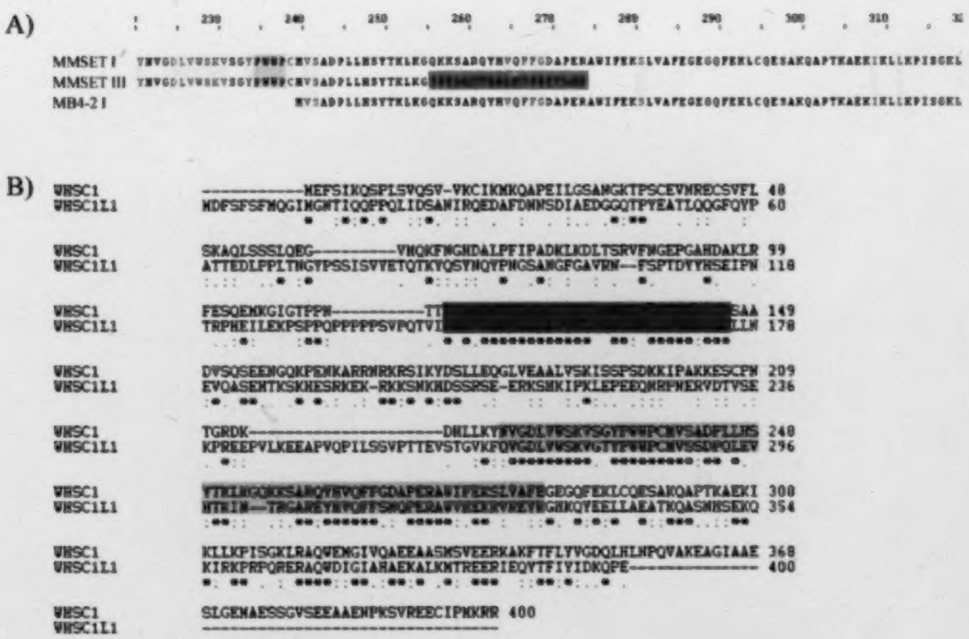
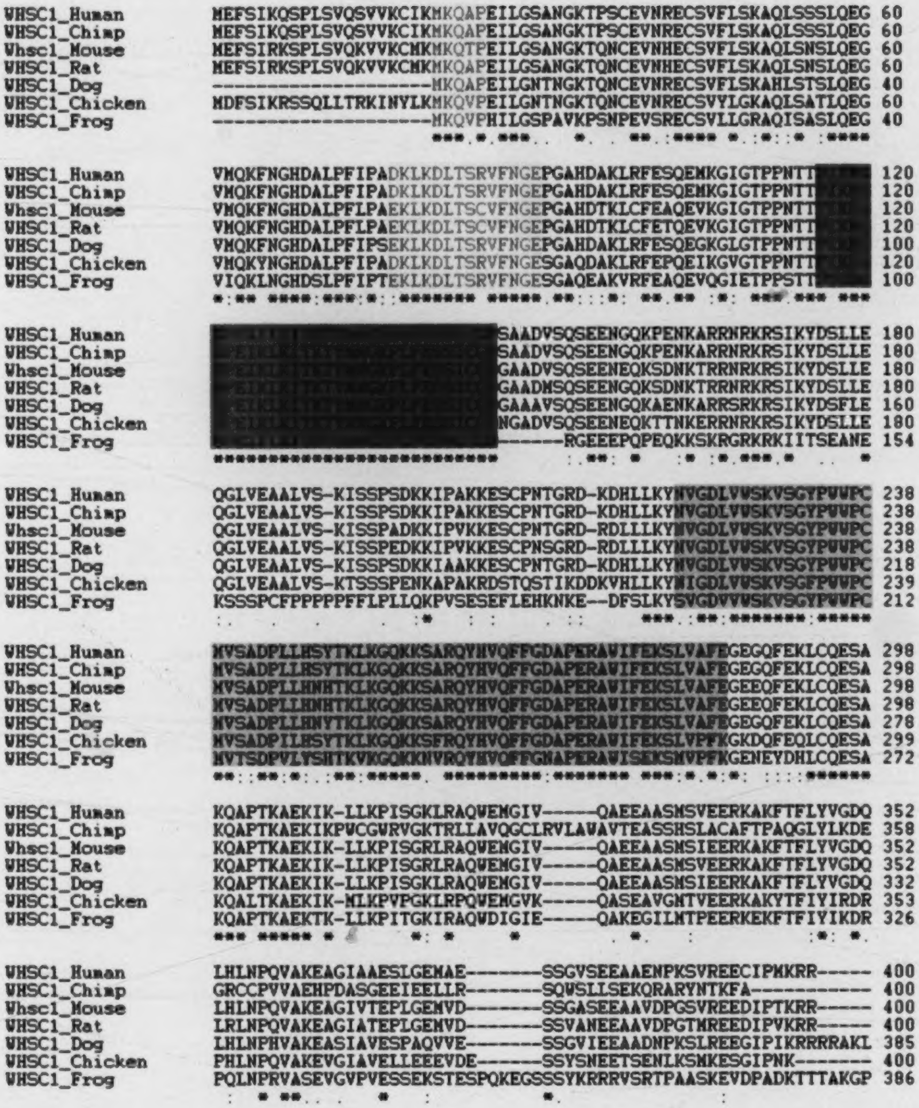


Figure C3.11 – N-Terminal Conservation between WHSC1/MMSET and WHSC1L1

A) The overlapping protein sequence of MMSET I, MMSET III, and MB4-2 I is shown. The highly conserved P-W-W-P motif is highlighted in red and additional residues conserved between various PWWP domains are highlighted in yellow[205]. The additional amino acids encoded by *MMSET* exon 4a in the MMSET III variant are highlighted in purple. Therefore, the highly conserved N-termini of the PWWP domain and the C-termini essential for the proper folding of the domain are lost in MB4-2 I and MMSET III variants respectively. And thus it is unlikely that the PWWP domain regulates the localization of wild-type MMSET variants. B) A pair-wise sequence alignment of the first 400 amino acids of WHSC1/MMSET and the closest human paralog WHSC1L1 is shown. The alignment was generated with the ClustalW sequence alignment tool (http://www.ebi.ac.uk/clustalw/) using default parameters. Identical amino acids are indicated by asterisks, conserved amino acid changes are indicated by colons, and semi-conserved changes are indicated by periods. The region encoding the PWWP domain identified by the SMART protein prediction program is highlighted in red and a novel region of high identity between amino acids 116-146 of WHSC1/MMSET is highlighted in green. The protein sequences used for this analysis are as follows: *Homo sapiens WHSC1* (NP_579890.1), *WHSC1L1* (NP_075447.1)

The alignment of the N-terminal 400 amino acids from WHSC1/MMSET and WHSC1L1 identified a region of high identity between amino acids 116-146 of the wild-type MMSET variants. To test if this region is conserved in evolution we aligned the N-terminal 400 amino acids of the identified WHSC1 orthologs (Figure C3.12). As expected the PWWP domain was conserved in all species. Not surprisingly the multiple sequence alignment identified a region of high identity C-terminal of the defined PWWP domain identified by the SMART protein prediction program. Therefore, as suggested by Qui et al. additional C-terminal amino acids are likely essential for the proper folding and function of this domain[205]. The entire region of high sequence identity between WHSC1/MMSET and WHSC1L1 is conserved in WHSC1 orthologs found in *Homo sapiens, Pan troglodytes, Canis familiaris, Mus musculus, Rattus norvegicus, Gallus gallus*, and *Xenopus tropicalis* (Figure C3.12). However, when organisms from the taxonomy class actinopterygii; *Tetraodon nigroviridis, Takifugu rubripes*, and *Danio rerio* are included only a small sub-region between amino acids 122-126 is highly conserved. Moreover, two additional regions of high identity, which are not shared with WHSC1L1, were identified between amino acids 21-25 and 77-90 of wild-type MMSET variants. Therefore, the region essential for the localization of wild-type MMSET proteins is likely encoded by one of these three sub-regions. A database search for potential functional sites in these evolutionarily conserved regions with the eukaryotic linear motif resource server (http://elm.eu.org/) identified five different motifs in the conserved region between amino acids 116-146 of wild-type MMSET variants. No predicted functional sites were identified between amino acids 21-25 or 77-90. Interestingly, only one functional site, the Class IV WW domain interaction motif, overlaps the evolutionarily conserved regions between amino acids 122-126. However, an essential residue in the motif is not conserved in *Danio rerio*, *Tetraodon nigroviridis*, and *Takifugu rubripes*. In the future it will be necessary to determine which of the evolutionarily conserved motifs are essential for the localization of wild-type MMSET variants.



```
WHSC1_Human      MEFSIKQSPLSVQSVVKCIKMKQAPEILGSANGKTPSCEVNRECSVFLSKAQLSSSLQEG 60
WHSC1_Chimp      MEFSIKQSPLSVQSVVKCIKMKQAPEILGSANGKTPSCEVNRECSVFLSKAQLSSSLQEG 60
Whsc1_Mouse      MEFSIRKSPLSVQKVVKCMKMKQTPEILGSANGKTQNCEVNHECSVFLSKAQLSNSLQEG 60
WHSC1_Rat        MEFSIRKSPLSVQKVVKCMKMKQAPEILGSANGKTQNCEVNHECSVFLSKAQLSNSLQEG 60
WHSC1_Dog        --------------------MKQAPEILGNTNGKTQNCEVNRECSVFLSKAHLSTSLQEG 40
WHSC1_Chicken    MDFSIKRSSQLLTRKINYLKMKQVPEILGNTNGKTQNCEVNRECSVYLGKAQLSATLQEG 60
WHSC1_Frog       --------------------MKQVPHILGSPAVKPSNPEVSRECSVLLGRAQISASLQEG 40
                                     ***.*.***..  *.  .**.:**** *.:*::* :****

WHSC1_Human      VMQKFNGHDALPFIPADKLKDLTSRVFNGEPGAHDAKLRFESQEMKGIGTPPNTT[illegible] 120
WHSC1_Chimp      VMQKFNGHDALPFIPADKLKDLTSRVFNGEPGAHDAKLRFESQEMKGIGTPPNTT[illegible] 120
Whsc1_Mouse      VMQKFNGHDALPFLPAEKLKDLTSCVFNGEPGAHDTKLCFEAQEVKGIGTPPNTT[illegible] 120
WHSC1_Rat        VMQKFNGHDALPFLPAEKLKDLTSCVFNGEPGAHDTKLCFETQEVKGIGTPPNTT[illegible] 120
WHSC1_Dog        VMQKFNGHDALPFIPSEKLKDLTSRVFNGEPGAHDAKLRFESQEGKGLGTPPNTT[illegible] 100
WHSC1_Chicken    VMQKYNGHDALPFIPADKLKDLTSRVFNGESGAQDAKLRFEPQEIKGVGTPPNTT[illegible] 120
WHSC1_Frog       VIQKLNGHDSLPFIPTEKLKDLTSRVFNGESGAQEAKVRFEAQEVQGIETPPSTT[illegible] 100
                 *:** ****:***:*::******* *****.**:::*: **.** :*: ***.*** ***

WHSC1_Human      [illegible]SAADVSQSEENGQKPENKARRNRKRSIKYDSLLE 180
WHSC1_Chimp      [illegible]SAADVSQSEENGQKPENKARRNRKRSIKYDSLLE 180
Whsc1_Mouse      [illegible]GAADVSQSEENEQKSDNKTRRNRKRSIKYDSLLE 180
WHSC1_Rat        [illegible]GAADMSQSEENGQKSDNKTRRNRKRSIKYDSLLE 180
WHSC1_Dog        [illegible]GAAAVSQSEENGQKAENKARRSRKRSIKYDSFLE 160
WHSC1_Chicken    [illegible]NGADVSQSEENEQKTTNKERRNRKRSIKYDSLLE 180
WHSC1_Frog       [illegible]------RGEEEPQPEQKKSKRGRKRKIITSEANE 154
                 **************************      :.**: *   :* :*.***.*  ..   *

WHSC1_Human      QGLVEAALVS-KISSPSDKKIPAKKESCPNTGRD-KDHLLKYNVGDLVWSKVSGYPWWPC 238
WHSC1_Chimp      QGLVEAALVS-KISSPSDKKIPAKKESCPNTGRD-KDHLLKYNVGDLVWSKVSGYPWWPC 238
Whsc1_Mouse      QGLVEAALVS-KISSPADKKIPVKKESCPNTGRD-RDLLLKYNVGDLVWSKVSGYPWWPC 238
WHSC1_Rat        QGLVEAALVS-KISSPEDKKIPVKKESCPNSGRD-RDLLLKYNVGDLVWSKVSGYPWWPC 238
WHSC1_Dog        QGLVEAALVS-KISSPSDKKIAAKKESCPNTGRD-KDHLLKYNVGDLVWSKVSGYPWWPC 218
WHSC1_Chicken    QGLVEAALVS-KTSSSPENKAPAKRDSTQSTIKDDKVHLLKYNIGDLVWSKVSGFPWWPC 239
WHSC1_Frog       KSSSPCFPPPPPFFLPLLQKPVSESEFLEHKNKE--DFSLKYSVGDVVWSKVSGYPWWPC 212
                 :.   .   .     .  :*   : :    .  ::     ***.:**:*******:*****

WHSC1_Human      MVSADPLLHSYTKLKGQKKSARQYHVQFFGDAPERAWIFEKSLVAFEGEGQFEKLCQESA 298
WHSC1_Chimp      MVSADPLLHSYTKLKGQKKSARQYHVQFFGDAPERAWIFEKSLVAFEGEGQFEKLCQESA 298
Whsc1_Mouse      MVSADPLLHNHTKLKGQKKSARQYHVQFFGDAPERAWIFEKSLVAFEGEEQFEKLCQESA 298
WHSC1_Rat        MVSADPLLHNHTKLKGQKKSARQYHVQFFGDAPERAWIFEKSLVAFEGEEQFEKLCQESA 298
WHSC1_Dog        MVSADPLLHNYTKLKGQKKSARQYHVQFFGDAPERAWIFEKSLVAFEGEGQFEKLCQESA 278
WHSC1_Chicken    MVSADPILHSYTKLKGQKKSFRQYHVQFFGDAPERAWIFEKSLVPFKGKDQFEQLCQESA 299
WHSC1_Frog       MVTSDPVLYSHTKVKGQKKNVRQYHVQFFGNAPERAWISEKSMVPFKGENEYDHLCQESA 272
                 **::**:*:. **:*****. *********:******* ***:*.*:*:  :::******

WHSC1_Human      KQAPTKAEKIK-LLKPISGKLRAQWEMGIV-----QAEEAASMSVEERKAKFTFLYVGDQ 352
WHSC1_Chimp      KQAPTKAEKIKPWCGWRVGKTRLLAVQGCLRVLAWAVTEASSHSLACAFTPAQGLYLKDE 358
Whsc1_Mouse      KQAPTKAEKIK-LLKPISGRLRAQWEMGIV-----QAEEAASMSIEERKAKFTFLYVGDQ 352
WHSC1_Rat        KQAPTKAEKIK-LLKPISGRLRAQWEMGIV-----QAEEAASMSVEERKAKFTFLYVGDQ 352
WHSC1_Dog        KQAPTKAEKIK-LLKPISGKLRAQWEMGIV-----QAEEAASMSIEERKAKFTFLYVGDQ 332
WHSC1_Chicken    KQALTKAEKIK-MLKPVPGKLRPQWEMGVK-----QASEAVGMTVEERKAKYTFIYIRDR 353
WHSC1_Frog       KQAPTKAEKTK-LLKPITGKIRAQWDIGIE-----QAKEGILMTPEERKEKFTFIYIKDR 326
                 *** ***** *       *: *     *        . *.   :          :*: *.

WHSC1_Human      LHLNPQVAKEAGIAAESLGEMAE--------SSGVSEEAAENPKSVREECIPMKRR----- 400
WHSC1_Chimp      GRCCPVVAEHPDASGEEIEELLR--------SQWSLLSEKQRARYNTKFA----------- 400
Whsc1_Mouse      LHLNPQVAKEAGIVTEPLGEMVD--------SSGASEEAAVDPGSVREEDIPTKRR----- 400
WHSC1_Rat        LRLNPQVAKEAGIATEPLGEMVD--------SSVANEEAAVDPGTMREEDIPVKRR----- 400
WHSC1_Dog        LHLNPHVAKEASIAVESPAQVVE--------SSGVIEEAADNPKSLREEGIPIKRRRRAKL 385
WHSC1_Chicken    PHLNPQVAKEVGIAVELLEEEVDE-------SSYSNEETSENLKSMKESGIPNK------- 400
WHSC1_Frog       PQLNPRVASEVGVPVESSEKSTESPQKEGSSSYKRRRVSRTPAASKEVDPADKTTTAKGP 386
                  :  * **.. .   *   :          *.              :
```

Figure C3.12 – Multiple Sequence Alignment of WHSC1/MMSET Orthologs

The multiple sequence alignment was generated with the ClustalW sequence alignment tool (http://www.ebi.ac.uk/clustalw/) using default parameters. Identical amino acids are indicated by asterisks, conserved amino acid changes are indicated by colons, and semi-conserved changes are indicated by periods. The region encoding the PWWP domain identified by the SMART protein prediction program is highlighted in red and the novel region of high identity between human WHSC1/MMSET and WHSC1L1 is highlighted in green. The N-terminal protein sequences from the WHSC1 orthologs in *Danio rerio*, *Takifugu rubripes*, and *Tetraodon nigroviridis* are not shown, however, the highly conserved regions shared by these organisms and those shown are highlighted in yellow. The gene identification tags and protein sequences retrieved from the NCBI and Ensembl databases for this analysis are as follows: *Homo sapiens WHSC1* (NP_579890.1); *Pan troglodytes* WHSC1/LOC461070 (XP_517068); *Canis familiaris* WHSC1/LOC479077 (XP_536224.1); *Mus musculus Whsc1* (XP_132006.4); *Rattus norvegicus* WHSC1/LOC305456 (XP_223540); *Gallus gallus* WHSC1/LOC422897 (XP_420839); *Xenopus tropicalis* WHSC1/ENSXETG00000002393 (ENSXETP00000005100); *Tetraodon nigroviridis* WHSC1/GSTENG00023330001 (GSTENP00023330001); *Takifugu rubripes* WHSC1/SINFRUG00000138891 (GENSCAN00000001541); *Danio rerio* WHSC1 (GENSCAN00000001643).

The full length MMSET variant, MMSET II, is predicted to be a histone methyltransferase as it contains a SET domain. Our observation of the strong co-localization and high affinity interaction between MMSET II and DNA/chromatin further supports the notion that this protein is a functional histone methyl-transferase. Moreover, the *WHSC1/MMSET* paralog *NSD1* was previously shown to be a functionally histone methyl-transferase specific for H3-K36 and H4-K20[200]. Therefore, MMSET is likely a functional histone methyl-transferase. To assess the potential histone methyl-transferase capability of MMSET II we compared the C-terminal sequence of MMSET, WHSC1L1, and NSD1 (Figure C3.13). To ensure that known essential amino acids are conserved in

MMSET and WHSC1L1 we mapped all of the NSD1 missense mutations associated with Sotos syndrome to the alignment, as mutation of these amino acids of NSD1 are predicted to inactivate the protein. In all cases the essential amino acids are conserved in MMSET. Surprisingly, several are not conserved in WHSC1L though none are within the SET domain. We then annotated the alignment with all of the residues conserved within the SET domain of proteins in the SET2 family of histone methyltransferases. Interestingly, five of the six missense mutations in the SET domain of NSD1 associated with Sotos syndrome are conserved in all SET2 family members and the sixth missense mutation occurs in a residue conserved in all NSD family members (Figure C3.13). Finally, the NHSC and GE$(X)_5$Y motifs essential for the function of the SET domain were mapped to the alignment[242]. MMSET and NSD1 share an identical GE$(X)_5$Y motif that does not influence the histone methyltransferase function of NSD1[200]. Surprisingly, the NHSC motif is different in these two paralogs as NSD1 contains a NHCC motif while MMSET contains the characteristic NHSC motif. Importantly, an engineered NSD1 mutant with an NHSC motif has a 20 fold higher methyltransferase activity compared to the wild-type NHCC motif[200]. Therefore, based on the sequence comparison, the SET domain in MMSET II is likely functional as it contains the NHSC motif, and all amino acids essential to the function of NSD1.

```
WHSC1           -----------SLK-SQAATKN--LSDACKP----------------LKKRNRASTAASS 609
WHSC1L1         -----------VKTGLQKGASE--ISDSCKP----------------LKKRSRASTDVEM 642
NSD1            GFMPKKGDLGLSKKCYEAGHLENGITESCATSYSKDFGGGTTKIFDKPRKRKRQRHAAAK 1480
                             .  :.   :  ::::* .                 :**.*    .

WHSC1           ALGFSKSSSPSASLTENEVSDSPGDEPSESPYESADETQTEVS-VSSKK---SERGV-TA 664
WHSC1L1         TSSAYRDTSDSDSRGLSDLQVGFG-KQVDSPSATADADVSDVQSMDSSL---SRRGTGMS 698
NSD1            MQCKKVKNDDSSKEIPGSEGELMPHRTATSPKETVEEGVEHDPGMPASKKMQGERGGGAA 1540
                      ... * .   ..       .   **  :.:    .   : :.    ..**   :

WHSC1           KKEYVCQLCEKPG-SLLLCEGPCCGAFHLACLGISRRPEGRFTCSECASGIHSCFVCKES 723
WHSC1L1         KKDTVCQICESSGDSLIPCEGECCKHFHLECLGLASLPDSKFICMECKTGQHPCFSCKVS 758
NSD1            LKENVCQNCEKLG-ELLLCEAQCCGAFHLECLGLTEMPRGKFICNECRTGIHTCFVCKQS 1599
                 *: *** **. * .*: **. **  *** ***:   * .:* * ** :* *.** ** *

WHSC1           KTDVKRCVVTQCGKFYHEACVKKYPLTVFESRGFRCPLHSCVSCHASNPSNPRPSKGKMM 783
WHSC1L1         GKDVKRCSVGACGKFYHEACVRKFPTAIFESKGFRCPQHCCSACSMEKDIH-KASKGRMM 817
NSD1            GEDVKRCLLPLCGKFYHEECVQKYPPTVMQNKGFRCSLHICITCHAANPANVSASKGRLM 1659
                  ***** :  ******* **:*:* ::::.:****. * * :*   :  :  .***::*

WHSC1           RCVRCPVAYHSGDACLAAGCSVIASNSIICTAHFTARKGKRHHAHVNVSWCFVCS----- 838
WHSC1L1         RCLRCPVAYHSGDACIAAGSMLVSSYILICSNHS---KRSSNSSAVNVGFCFVCARGLIV 874
NSD1            RCVRCPVAYHANDFCLAAGSKILASNSIICPNHFTPRRGCRNHEHVNVSWCFVCS----- 1714
                **:*******:.* *:***. :::*  :**. *    :   :   ***.:****:

WHSC1           --------------------------------------------KGGSLLCCESCPAAFH 854
WHSC1L1         QDHSDPMFSSYAYKSHYLLNESNRAELMKLPMIPSSSASKKKCEKGGRLLCCESCPASFH 934
NSD1            --------------------------------------------EGGSLLCCDSCPAAFH 1730
                                                            :** ****:****:**

WHSC1           PDCLNIEMPDGSWFCNDCRAGKKLHFQDIIWVKLGNYRWWPAEVCHPKNVPPNIQKMKHE 914
WHSC1L1         PECLSIEMPEGCWNCNDCKAGKKLHYKQIVWVKLGNYRWWPAEICNPRSVPLNIQGLKHD 994
NSD1            RECLNIDIPEGNWYCNDCKAGKKPHYREIVWVKVGRYRWWPAEICHPRAVPSNIDKMRHD 1790
                 :**.*::*:* * ****:**** *:::*:***:*.*******:*:*: ** **: ::*:

WHSC1           IGEFPVFFFGSKDYYWTHQARVFPYMEGDRGSRYQGVRGIGRVFKNALQEAEARFREIKL 974
WHSC1L1         LGDFPVFFFGSHDYYWVHQGRVFPYVEGDK-SFAEGQTSINKTFKKALEEAAKRFQELKA 1053
NSD1            VGEFPVLFFGSNDYLWTHQARVFPYMEGDVSSKDKMGKGVDGTYKKALQEAAARFEELKA 1850
                :*:***:****:** *.**.*****:***  *  :   .:. .:*:**:**  **.*:*

WHSC1           QREARETQESE---RKPPPYKHIKVNKPYGKVQIYTADISEIPKCNCKPTDENPCGFDSE 1031
WHSC1L1         QRESKEALEIEKNSRKPPPYKHIKANKVIGKVQIQVADLSEIPRCNCKPADENPCGLESE 1113
NSD1            QKELRQLQEDRKNDKKPPPYKHIKVNRPIGRVQIFTADISEIPRCNCKATDENPCGIDSE 1910
                *:* ::  * .   :*********.*:  *:*** .**:****:****.:******::**

WHSC1           CLNRMLMFECHPQVCPAGEFCQNQCFTKRQYPETKIIKTDGKGWGLVAKRDIRKGEFVNE 1091
WHSC1L1         CLNRMLQYECHPQVCPAGDRCQNQCFTKRLYPDAEIIKTERRGWGLRTKRSIKKGEFVNE 1173
NSD1            CINRMLLYECHPTVCPAGGRCQNQCFSKRQYPEVEIFRTLQRGWGLRTKTDIKKGEFVNE 1970
                *:**** :**** *****  ******:** **:.:*::*  :**** :* .*:*******

WHSC1           YVGELIDEEECMARIKHAHENDITHFYMLTIDKDRIIDAGPKGNYSRFM[illegible]QPNCETL 1151
WHSC1L1         YVGELIDEEECRLRIKRAHENSVTNFYMLTVTKDRIIDAGPKGNYSRFM[illegible]NPNCETQ 1233
NSD1            YVGELIDEEECRARIRYAQEHDITNFYMLTLDKDRIIDAGPKGNYARFM[illegible]QPNCETQ 2030
                ***********  **: *:*:.:*:*****: *************:*****.*:*****

WHSC1           KWTVNGDTRVGLFAVCDIPA[illegible]NLDCLGNEKTVCRCGASNCSGFLGDRPKTSTT 1211
WHSC1L1         KWTVNGDVRVGLFALCDIPA[illegible]NLDCLGNGRTECHCGADNCSGFLGVRPKSACA 1293
NSD1            KWSVNGDTRVGLFALSDIKA[illegible]NLECLGNGKTVCKCGAPNCSGFLGVRPKNQPI 2090
                **:****.******:.** ** ********:**** :* *:*** ******* ***.

WHSC1           LSSEEKGKKTKKKTRRRRAKGEGKRQSEDECFRCGDGGQLVLCDRKFCTKAYHLSCLGLG 1271
WHSC1L1         STNEEKAKNAKLKQKRRKIKTEPKQMHEDYCFQCGDGGELVMCDKKDCPKAYHLLCLNLT 1353
NSD1            ATEEKSKKFKKKQQGKRRTQGEITKEREDECFSCGDAGQLVSCKKPGCPKVYHADCLNLT 2150
                 :.*:. *  * :  :*: : * .:  ** ** ***.*:** *.:  *.*.**  **.*

WHSC1           KRPFGKWECPWHHCDVCGKPSTSFCHLCPNSFCKEHQDGTAFSCTPDGRSYCCEHDLGAA 1331
WHSC1L1         QPPYGKWECPWHQCDECSSAAVSFCEFCPHSFCKDHEKGALVPSALEGRLCCSEHDPMAP 1413
NSD1            KRPAGKWECPWHQCDICGKEAASFCEMCPSSFCKQHREGMLFISKLDGRLSCTEHDPCGP 2210
                : * ********:** *.. :.***.:** ****:*..*  . .  :**  * ***  ..

WHSC1           S---VRSTKTEKPPPE--PGKPKGKRRRRRGWRRVTEGK--------------------- 1365
WHSC1L1         -----------VSPEY--WSKIKCKWESQDHGEEVKE----------------------- 1437
NSD1            NPLEPGEIREYVPPPVPLPPGPSTHLAEQSTGMAAQAPKMSDKPPADTNQMLSLSKKALA 2270
                            .*        . :   :     .
```

Figure C3.13 – Multiple Sequence Alignment of the C-Terminus of Human NSD Paralogs

The multiple sequence alignment was generated with the ClustalW sequence alignment tool (http://www.ebi.ac.uk/clustalw/) using default parameters. Identical amino acids are indicated by asterisks, conserved amino acid changes are indicated by colons, and semi-conserved changes are indicated by periods. The regions encoding the PHD, PWWP, and SET (AWS, SET, post SET) domains identified by the SMART protein prediction program are indicated by blue, red and gold bars above the alignments, respectively. The missense mutations of NSD1 recorded in the Human Gene Mutation Database (http://www.hgmd.org/), which are associated with Sotos syndrome, are highlighted in red. Amino acids conserved in the SET (AWS, SET, post SET) domains of all SET2 histone methyltransferase family members are indicated by red asterisks while amino acids conserved in all NSD family members are indicated by blue asterisks. The protein sequences used for this analysis are as follows: *Saccharomyces cerevisiae* SET1 (NP_011987.1); *Drosophila melanogaster* Mes-4/CG4976-PA (NP_733239.1), Ash1/CG8887-PA (NP_524160.1), HYPB, like/CG1716-PA (NP_572888); *Homo sapiens HYPB* (NP_054878.3), *ASH1L* (NP_060959.1), *WHSC1* (NP_579890.1), *WHSC1L1* (NP_075447.1), *NSD1* (NP_071900.2); *Mus musculus Whsc1* (XP_132006.4); *Xenopus tropicalis* WHSC1/ENSXETG00000002393 (ENSXETP00000005100); *Gallus gallus* WHSC1/LOC422897 (XP_420839); *Danio rerio* WHSC1 (GENSCAN00000001643)

C3.4.1 – Chapter Conclusions

The principle goal of this chapter was to determine which overexpressed transcript from the MMSET locus is most likely to contribute to t(4;14) myelomagenesis. We previously identified three different and universally dysregulated transcripts in all t(4;14) positive patients. However, the overexpressed transcripts encoding MMSET I or MMSET II are not the same in all t(4;14) positive patients, since patients with MB4-2 and MB4-3 breakpoints do not contain the wild-type translation initiation site. Therefore, the only universally dysregulated transcript encoding an identical protein product in all t(4;14) positive patients is the transcript encoding RE-IIBP. However, it was possible that the overexpressed transcripts in patients with MB4-2 and MB4-3 breakpoints produced truncated protein products from downstream alternative translation initiation sites. Moreover, even though these variant proteins are N-terminally truncated it was possible that they would be functionally equivalent to the wild-type MMSET variants. Therefore we analyzed the protein products of the various transcripts to determine if they fit our assumption that the myelomagenic protein product would be functionally equivalent in all t(4;14) positive patients regardless of breakpoint type, since the ability to encoded wild-type or truncated MMSET variants does not influence survival (See C2.3.4).

Following this assumption we were able to prove our working hypothesis that the only transcript from the MMSET locus overexpressed as a result of t(4;14) and producing a protein with uncompromised function is the RE-IIBP encoding transcript. We found the wild-type MMSET I and MMSET II proteins were localized to the nucleus and excluded from nucleoli while unexpectedly, the majority of RE-IIBP was located in cytoplasmic foci with a secondary pool in nucleoli. Using anti-GFP immunoblots we showed that the overexpressed transcripts encoding MMSET I and MMSET II from MB4-2 and MB4-3 patients did produce truncated protein products. The molecular weight of the observed bands was comparable to the predicted molecular weight of proteins originating from the alternative translation initiation sites in *MMSET* exon 4 or 6. However, these truncated MMSET variants have a different localization pattern compared to the wild-type MMSET variants. They localize to the nucleus and appear to be enriched in nucleoli. Though the different localization patterns suggested a loss of function in the truncated variants, it is still possible that the truncated variants are functionally equivalent to the wild-type variants. Unfortunately, the function of MMSET has not yet been elucidated, so we indirectly tested the function of the different MMSET variants with FRAP. Using this kinetic measurement of protein function it is evident that the truncated MMSET variants, from MB4-2 and MB4-3 patients, are not functionally equivalent to the wild-type variants, produced in patients with MB4-1 breakpoints. Therefore, the only overexpressed transcript from the MMSET locus in t(4;14) positive myeloma with and uncompromised protein function in all patients is the transcript encoding RE-IIBP.

The different localization and kinetic profiles suggested an essential domain for the localization and function of MMSET exists in the N-terminus. To determine if the N-terminus regulated the localization of MMSET, we cloned a natural MMSET splice variant transcript encoding a protein corresponding to the protein sequence lost in MB4-2 variants. This protein variant, MMSET III, localized to the nucleus and was excluded from nucleoli, equivalent to the wild-type protein variants. Therefore, the localization of wild-type MMSET variants is mediated by the N-terminal portion encoded by *MMSET* exons 3 and 4. Sequence analysis of this N-terminal region identified several evolutionarily conserved regions which are likely involved in regulating the localization of wild-type MMSET variants.

Chapter 4

Analysis of t(4;14) Genomic Breakpoints

C4.1.1 - Brief Introduction

Translocations involving the IgH locus are one of the most common genetic abnormalities in multiple myeloma[65]. However, unlike several other hematological malignancies, the translocations involve multiple partner chromosomes. In multiple myeloma the most common translocations, t(11;14)(q13;q32) and t(4;14)(p16;q32), are detectable in approximately 20% and 15% of patients, respectively[26,65].

Though translocations involving the IgH locus are common in other B-cell malignancies the molecular anatomy of the translocations in myeloma are different. Unlike the majority of IgH translocations observed in other B-cell malignancies, the translocations in multiple myeloma usually involve breakpoints within the switch regions of the IgH locus[232]. Analysis of the breakpoint junctions from t(4;14) samples suggested that these translocations result from illegitimate class switch recombination (CSR) events[88,96,232,243]. In general the der(4) breakpoint contained an Sμ-chromosome 4 junction and the der(14) contained a chromosome 4-downstream switch region junction. This fits with a model in which the CSR machinery has aberrantly joined chromosome 4 sequences to the switch regions involved in CSR. Therefore, the IgH translocations in multiple myeloma are typically called switch translocations to indicate this proposed mechanism. However, not all t(4;14) breakpoints fit this model of an illegitimate CSR event, since derivative chromosomes were observed with hybrid switch regions[89,232,236]. Several models can explain this observation, but in the only sample where both derivatives are cloned, the translocation mechanism does not appear to be linked to the class switch recombination process[89].

Although the translocation mechanism is assumed to involve aberrant CSR events, no functional evidence exists to validate this assumption. To prove this hypothesis, it is necessary to demonstrate a requirement for the CSR machinery in the recombination process. Alternatively, if the mechanism is related to CSR, the involved sequence regions should support class switch recombination. This type of experiment is usually performed in transgenic mouse models[244,245]; however, the entire breakpoint region (~65 kb) is too large to be independently tested. Therefore, before definitive experiments can be performed, more specific breakpoint regions must be defined. Furthermore, by cloning both derivatives from a large series of t(4;14) positive patients, we should be able to determine if IgH translocations are caused by aberrant CSR events or alternative mechanisms.

C4.2.1 – Specific Aim

1) *To clone both t(4;14) derivative chromosomes from a large series of t(4;14) positive patients to determine the potential translocation mechanism*

2) *To determine if recurrent breakpoint sites exist in chromosome 4p16*

C4.3 – Chapter 4 Results

All of the data in this chapter is currently unpublished; however, we are in the process of finishing the project and currently plan to submit this body of work for publication in the near future.

C4.3.1 – The Breakpoint Cloning Strategy

The der(4) IgH-MMSET and der(14) MMSET-IgH hybrid transcript assays identify three major breakpoint groups based on the observed product sizes. These breakpoint groups define the relative location of the breakpoint on chromosome 4. Based on the most centromeric and telomeric breakpoints cloned to date, the breakpoint region on chromosome 4 spans 64.5 kb between *LETM1* intron 2 and *MMSET* intron 4a. Patients in the MB4-1 group are predicted to have breakpoints telomeric of *MMSET* exon 3, while MB4-2 and MB4-3 patients are predicted to have breakpoints in *MMSET* intron 3 and 4/4a, respectively. Though the predicted breakpoint region for MB4-2 patients is only 3.0 kb, the predicted regions for MB4-1 and MB4-3 patients are relatively large, ~51.9 kb (based on most telomeric breakpoint cloned to date) and 12.5 kb, respectively. However, the predicted breakpoint regions for MB4-1 and MB4-3 patients can be refined to smaller sub-regions based on the hybrid transcript results. The MB4-3 patients can be separated into MB4-3a and MB4-3b subclasses based on the presence or absence of *MMSET* exon 4a in the der(4) IgH-MMSET hybrid transcript assays. This defines two separate regions of 7.7 kb (*MMSET* intron 4) and 4.7 kb (*MMSET* intron 4a), respectively. The MB4-1 patients can be separated into MB4-1a and MB4-1b subclasses based on the detection of der(14) MMSET-IgH hybrid transcripts. Patients with MB4-1 breakpoints and detectable der(14) MMSET-IgH hybrid transcripts must have breakpoints between *MMSET* exon 1 and 3, otherwise the der(14) reactions would fail to detect a product since the reactions use a primer in *MMSET* exon 1. Therefore, MB4-1a patients are predicted to have breakpoints in an approximate 22.8 kb region telomeric of *MMSET* exon 1 while MB4-1b patients are predicted to have breakpoints in a 29.1 kb region between MMSET exon 1 and 3. With these subclasses in mind, we designed a series of bi-directional PCR primers covering the approximate 64.5 kb breakpoint region on chromosome 4 at 5 kb intervals (Figure C4.1). Finally, by their very nature the hybrid transcript assays define the strand orientations of the IgH elements and the MMSET elements. Therefore, for example, in an Iμ1-ms6r positive patient with an MB4-2 breakpoint, we would expect the only limiting factor to a genomic Iμ1-ME4 PCR reaction to be the physical distance.

Since our previous work suggested the der(14) chromosome is lost in some t(4;14) positive patients, we attempted to amplify the der(4) breakpoint first. In general we used the Iμ1 primer as the 5' primer, as all but three of our patients have detectable der(4) hybrid transcripts with an Imu primer, and a series of 3' primers appropriate to each breakpoint subclass (Figure C4.1). The der(14) breakpoints were amplified with several issues in mind. First, a 3' primer was selected based on the IgH constant region associated with the der(14) MMSET-IgH hybrid transcripts. In patients with undetectable der(14) MMSET-IgH hybrid transcripts, the 3' primer was selected based on the clinical isotype of the patient, since the majority of the detected der(14) IgH-MMSET hybrid transcripts are associated with the same IgH constant region (See C2.3.3). As a last resort, primers specific to the mu, gamma, and alpha constant regions were used in parallel. The 5' primers were selected based on the der(4) IgH-MMSET hybrid transcript results and the cloned der(4) genomic breakpoint (Figure C4.1).

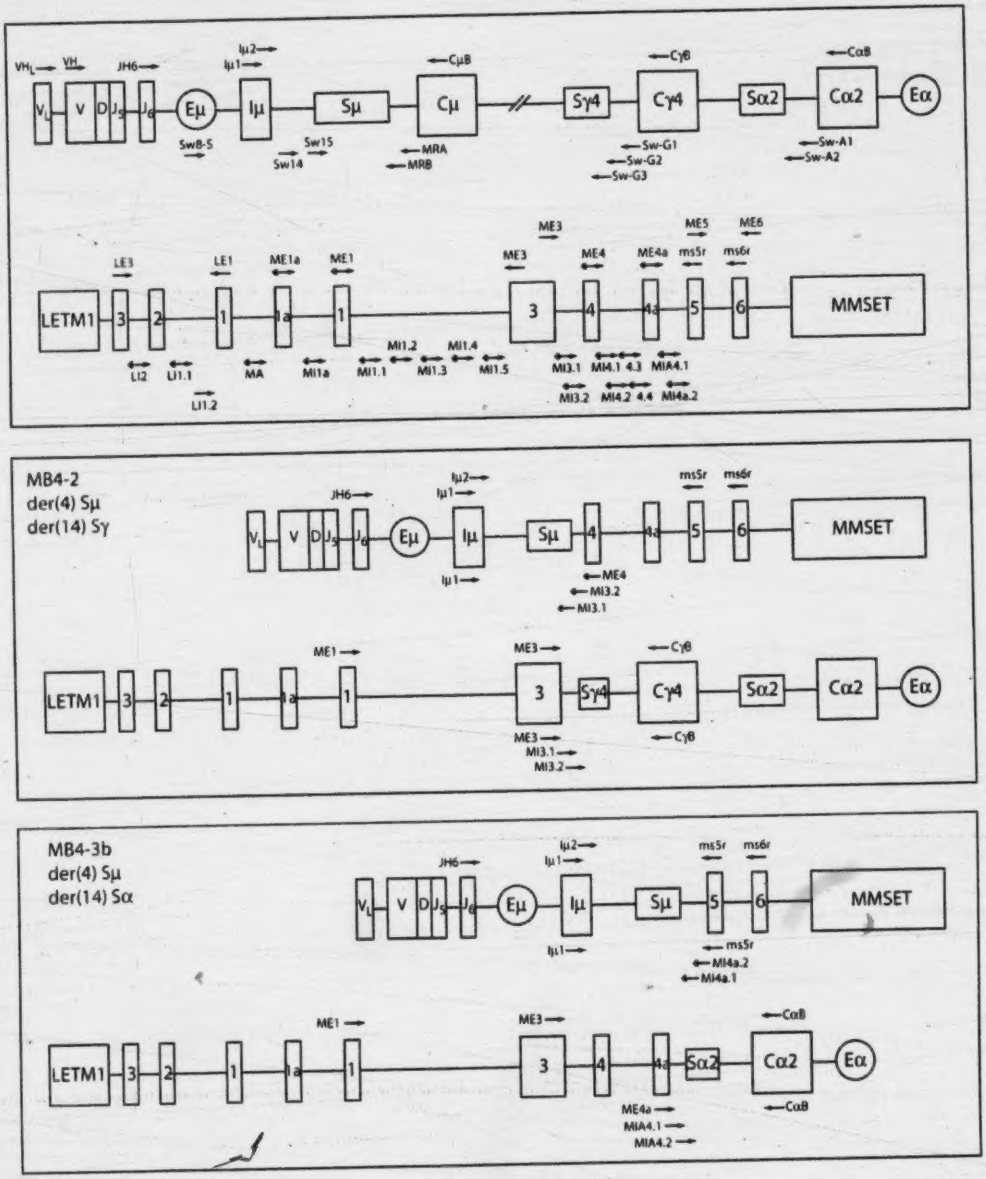

VH$_L$
VH
JH6
Iμ2
Iμ1
CμB
Cγ8
Cα8
V$_L$
V
D
J$_s$
J$_G$
Eμ
Iμ
Sμ
Cμ
Sγ4
Cγ4
Sα2
Cα2
Eα
Sw8-S
Sw15
Sw14
MRA
MRB
Sw-G1
Sw-G2
Sw-G3
Sw-A1
Sw-A2
ME3
ME5
ME6
LE3
LE1
ME1a
ME1
ME3
ME4
ME4a
ms5r
ms6r
LETM1
3
2
1
1a
1
3
4
4a
5
6
MMSET
MI1.2
MI1.4
LI2
LI1.1
MA
MI1a
MI1.1
MI1.3
MI1.5
MI3.1
MI4.1
4.3
MIA4.1
MI3.2
MI4.2
4.4
MI4a.2
LI1.2
MB4-2
der(4) Sμ
der(14) Sγ
Iμ2
Iμ1
JH6
ms5r
ms6r
Iμ1
ME4
MI3.2
MI3.1
ME1
ME3
Cγ8
ME3
MI3.1
MI3.2
Cγ8
MB4-3b
der(4) Sμ
der(14) Sα
Iμ2
Iμ1
JH6
ms5r
ms6r
Iμ1
ms5r
MI4a.2
MI4a.1
ME1
ME3
Cα8
ME4a
MIA4.1
MIA4.2
Cα8

Figure C4.1 – The Breakpoint Cloning Strategy

A general outline of the PCR based cloning strategy is shown. The upper panel illustrates the involved regions of 14q32 and 4p16. The relative position of primers used to detect the der(4) and der(14) hybrid transcripts along with the series of primers designed to amplify and sequence the genomic breakpoints are shown. The middle panel illustrates the predicted derivative chromosomes in a patient with MB4-2 der(4) IgH-MMSET hybrid transcripts and a detectable der(14) hybrid transcript involving an IgA locus. The primers used to detect the hybrid transcripts are shown above each respective exon and the primers used to amplify the genomic breakpoints are shown below each illustration. The lower panel illustrates the predicted derivative chromosomes in a patient with MB4-3b der(4) IgH-MMSET hybrid transcripts and a detectable der(14) hybrid transcript involving an IgA locus.

C4.3.2 – The Cloned t(4;14)(p16;q32) Breakpoints

We attempted to clone the der(4) genomic breakpoint from our t(4;14) positive patients and selected cell lines with MB4-2 and MB4-3 breakpoints using our PCR based strategy. Bone marrow DNA samples were available for 9 of the 15 patients identified to date with MB4-2 or MB4-3 breakpoints (See Appendix I). Using our PCR based strategy; we successfully amplified and sequenced the breakpoint from 7 of the 9 patients and 3 of 4 cell lines (Table C4.1). Surprisingly, the amplification strategy failed in KMS-18 even though the genomic breakpoint was previously described by Ronchetti et al.[243]. The reason for this failure is not known. Based on the der(4) IgH-MMSET hybrid assay we know the Imu region and *MMSET* exon 4 are present on the same chromosome and in the same transcriptional orientation. Potentially the distance between these two regions is two large or complex to amplify, though we have successfully amplified products over 12 kb with this strategy. Interestingly, the limited amount of sequence published by Ronchetti et al. aligns to multiple locations on both chromosomes. However, one of the multiple alignment matches on chromosome 4 ends 22 bp from the der(14) breakpoint that we cloned, suggesting this is the der(4) breakpoint site. The two patients in which the amplification strategy failed are unusual patients. Patient 1223 (described in C1.3.5) has detectable MB4-2 and MB4-4 der(4) IgH-MMSET hybrid transcripts. Patient 898 is the only FGFR3 non-expressor tested who has a detectable Iμ1-ms6r der(4) hybrid transcript product. In both cases the failed amplification attempts may reflect the occurrence of a complex rearrangement or the occurrence of breakpoints telomeric of *MMSET* with a subsequent internal deletion on chromosome 4 as seen in the KMS-28BM cell line[236].

Based on the cloned der(4) breakpoints we successfully amplified and sequenced the der(14) breakpoint from 5 of the 9 patients and 3 of the 3 cell lines attempted (Table C4.1). The failure of the amplification strategy in the four patients is not completely unexpected. In the case of patients, 1223 and 898, we also failed to amplify the der(4) breakpoint, and we expected the der(14) cloning reactions to fail in patients 898 and 1308, since they do not express FGFR3 or have detectable der(14) MMSET-IgH hybrid transcripts, suggesting that der(14) has been lost. The remaining patient, 1704, is one of our rare MB4-2 or MB4-3 patients expressing FGFR3 but with no detectable der(14) MMSET-IgH hybrid transcript (See C2.3.3). This may in part explain our inability to amplify the der(14) breakpoint from this patient. Never the less, we were able to successfully amplify the der(14) from the NCI-H929 cell line which has the same features of FGFR3 expression but undetectable der(14) MMSET-IgH hybrid transcripts.

Table C4.1 – Attempted Breakpoint Cloning Results

Patient or Cell Line	Breakpoint Type	FGFR3 Expression	der(14) Result	Clinical Isotype	der(4) Cloning	der(14) Cloning
Edm 1174	MB4-2	Positive	Detected, IgG	IgG	Successful	Successful
Edm 1850	MB4-2	Positive	Detected, IgA	Lambda	Successful	Successful
Edm 1223†	MB4-2	Positive	ND	IgG	Failed	Failed
Edm 898	MB4-2	Negative	ND	IgG	Failed	Failed
Edm 434	MB4-3a	Positive	Detected, IgG	IgG	Successful	Successful
Edm 1661	MB4-3a	Positive	Detected, IgG	IgG	Successful	Successful
Edm 1308‡	MB4-3a	Negative	ND	IgA	Successful	Failed
Edm 1704	MB4-3a	Positive	ND	IgG	Successful	Failed
Edm 1394	MB4-3a	Positive	Detected, IgA	IgG	Successful	Successful
NCI-H929	MB4-2	Positive	Negative	IgA	Successful*	Successful
KMS-18	MB4-2	Positive	Detected, IgG	IgA	Failed*	Successful
KMS-26	MB4-3a	Positive	Detected, IgG	IgG	Successful	Successful
OPM-2	MB4-3b	Positive	Negative	IgG	Successful	Not Done*

We attempted to clone the der(4) and der(14) breakpoints from patients and selected control cell lines with MB4-2, MB4-3a, and MB4-3b breakpoints identified by the der(4) IgH-MMSET hybrid transcript assays. We assessed all patients whose cancer had these breakpoints and for whom a bone marrow DNA sample was available. The breakpoint type, FGFR3 expression status, der(14) MMSET-IgH hybrid transcript result, and clinical isotype of each patient is provided.

† - Patient 1223 is the unique patient described in C1.3.5 with detectable MB4-2 and MB4-4 hybrid transcripts. Unsuccessful attempts were made to clone the potential der(4) breakpoints within MMSET intron 3 and 5, respectively.

‡ - Patient 1308 is one of the unique patients described in C1.3.5 with predicted breakpoints between the JH and Imu regions of the IgH locus.

ND – Not Detected.

* - Breakpoints previously cloned by other groups

C4.3.3 – Modeling the Cloned t(4;14)(p16;q32) Breakpoints

The IgH translocations observed in multiple myeloma are predicted to result from illegitimate class switch recombination events[73,232]. As discussed previously, although no direct evidence exists to link the translocations to the class switch recombination process the majority of the previously cloned breakpoints in t(4;14) and other IgH translocations in myeloma support this hypothesis, as the chromosome 14 breakpoint is generally in a switch region. Several models have been proposed based on the breakpoints cloned to date.

In the classic "primary switch" model, an illegitimate switch recombination occurs when an initial CSR event between switch mu and a downstream switch region aberrantly rejoins with chromosome 4 (Figure C4.2a)[88,96,232]. In this model switch mu is joined to chromosome 4 to form the der(4) and the downstream switch region is joined to chromosome 4 to form the der(14). However, several of the initially cloned breakpoints did not fit the "primary switch" model as they contained hybrid switch regions joined to chromosome 4.

This could be explained by the "secondary switch" model in which the illegitimate switch recombination occurs during a second CSR event in a cell that had previously undergone a successful isotype switching event (Figure C4.2b). However, when we consider the myeloma clone for those patients from whom both derivatives were cloned, not a single patient fits this model (Figure C4.3).

Two alternative non-classical models could potentially explain this discrepancy. The first model is the "CSR refinement" model proposed and functionally described by Kovalchuk et al. (Figure C4.2c)[246]. In this model a translocation event occurs in a single switch region and a subsequent CSR event refines a derivative to bring the IgH regulatory elements into closer proximity of the target gene. This model predicts the existence of a clonally heterogeneous tumour with CSR refined subclones likely having a selective advantage. However, if this model is correct, it must be possible for the refinement process to occur on both derivatives if the model is to fit with sequence data for the patients from which both derivatives are cloned (Figure C4.3). Moreover, beyond the involvement of a switch region, there is no reason to believe the initial translocation event involves the CSR machinery in the "CSR refinement" model. If the initial translocation event were linked to CSR, at a minimum, we would expect to see a deletion in the involved switch region. Although some of the patients whose myeloma breakpoints fit this model have small deletions or inversions within the involved switch region, others like OPM-2 have a perfectly conserved break within the involved switch region.

A second alternative model that we are calling the "post-switch strand invasion" model was proposed by Fenton et al. [89](Figure C4.2d). This model implies that a non-CSR mediated translocation event occurs within a previously rearranged hybrid switch junction. The strength of this model is that the break can occur in either switch mu or the rearranged downstream switch region and does not require a refinement event on either derivative, unlike the "CSR refinement" model. Furthermore, just like the "CSR refinement" model, all of the patients we characterized as having non-classical translocation events would fit the model (Figure C4.3).

Therefore, patients can be divided into those with classical "primary switch" translocations or those with non-classical "CSR refinement or post-switch strand invasion" translocations. Among patients for whom both derivatives were cloned, the frequency of each type of translocation is nearly equal, 46.2% versus 53.8%, respectively (Table C4.2). When we included patients with only a single cloned derivative the frequencies remain similar, 53.6% versus 46.4%, respectively (Table C4.2). The inversion of the relative frequencies may reflect either the true incidence, or the incorrect classification of several samples into the classical "primary switch" translocation group. In contrast, a predicted non-classical translocation breakpoint is unlikely to be misclassified because the presence of switch mu or a downstream switch region on the der(14) or der(4), respectively, clearly identifies patients fitting the non-classical models. Therefore, roughly half of the cloned t(4;14) translocations result from a classical "primary switch" translocation event and the other half result from non-classical translocation events.

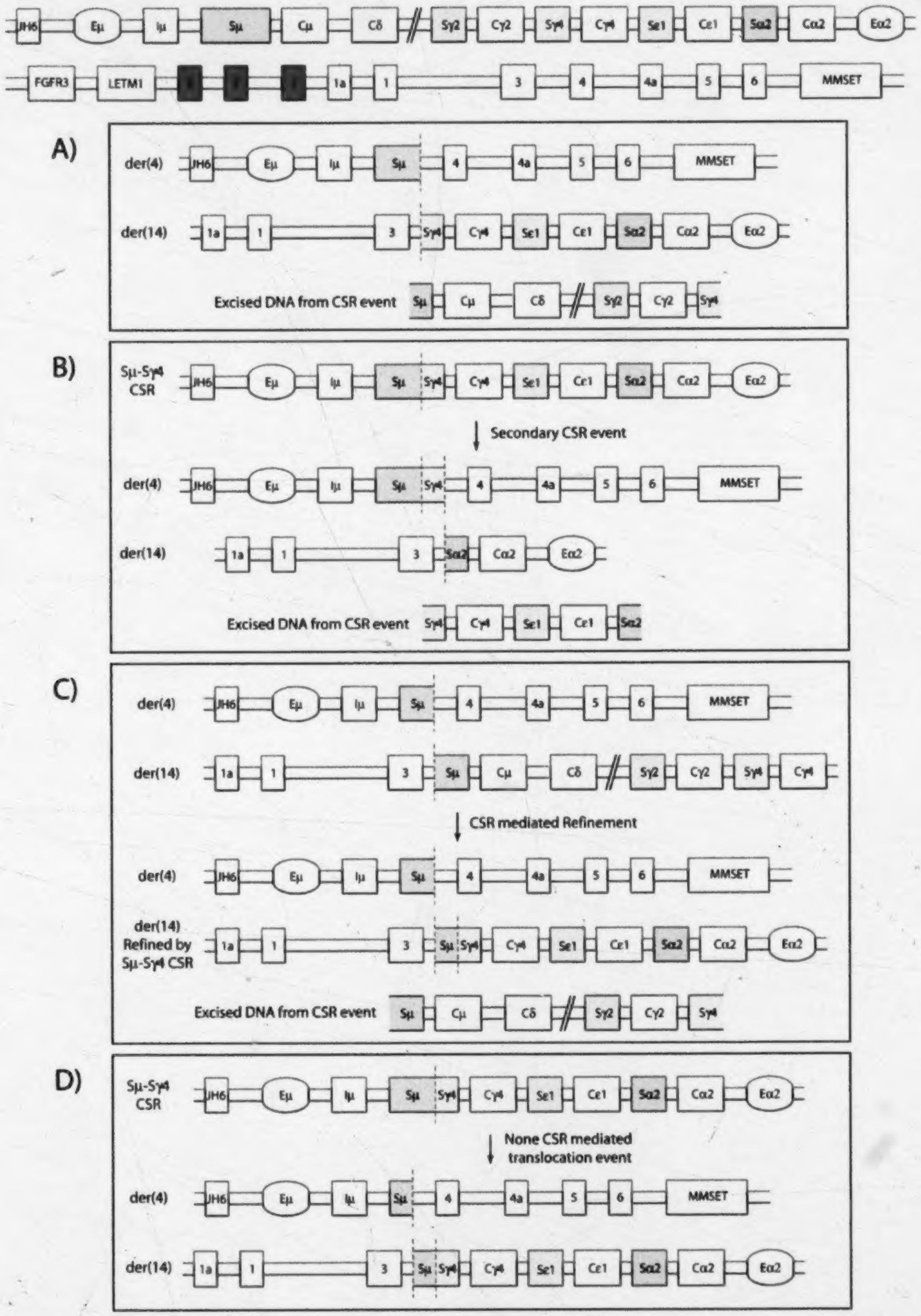
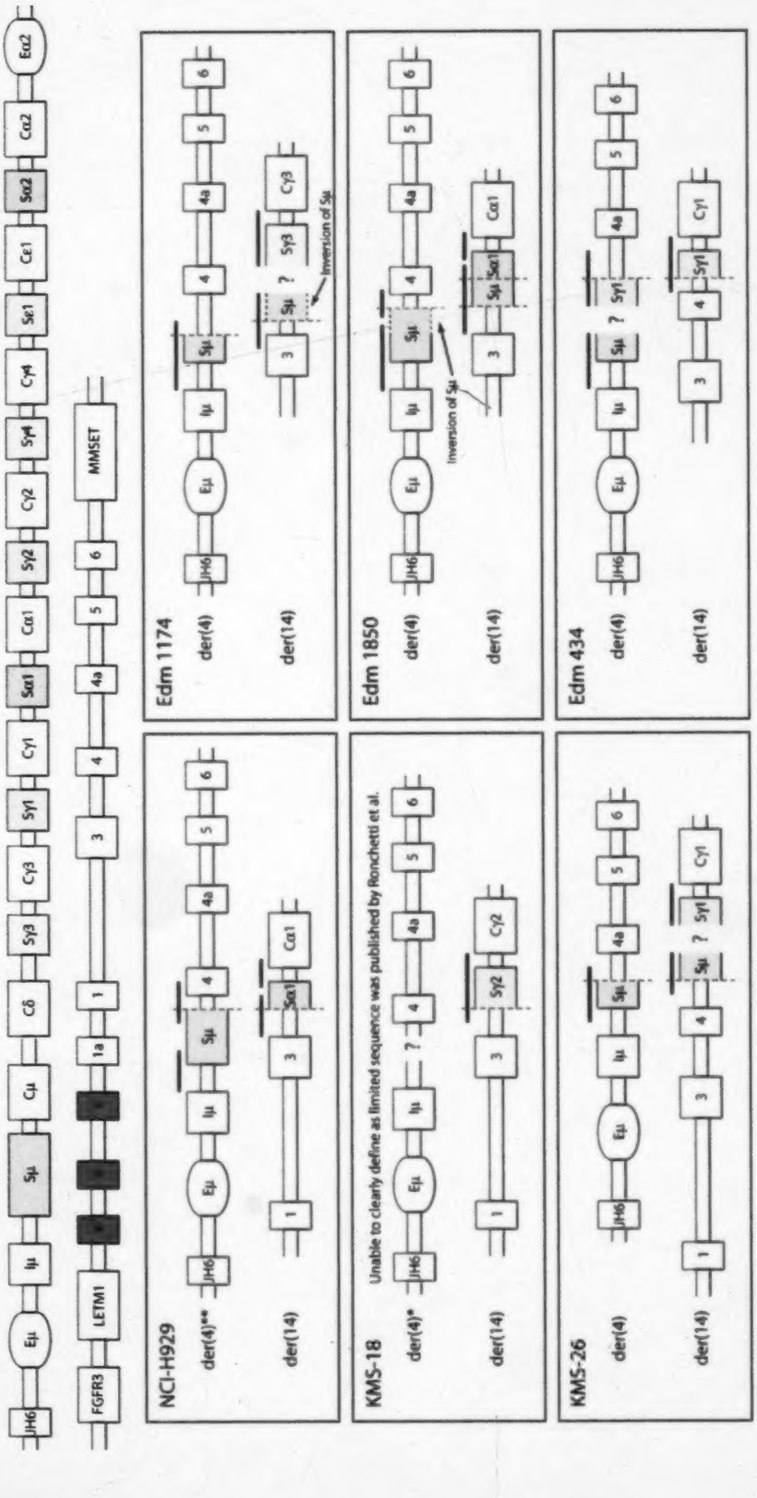

JH6
Eμ
Iμ
Sμ
Cμ
Cδ
Sγ2
Cγ2
Sγ4
Cγ4
Sε1
Cε1
Sα2
Cα2
Eα2
FGFR3
LETM1
1a
1
3
4
4a
5
6
MMSET
A)
der(4)
der(14)
Excised DNA from CSR event
B)
Sμ-Sγ4
CSR
Secondary CSR event
der(4)
der(14)
Excised DNA from CSR event
C)
der(4)
der(14)
CSR mediated Refinement
der(4)
der(14)
Refined by
Sμ-Sγ4 CSR
Excised DNA from CSR event
D)
Sμ-Sγ4
CSR
None CSR mediated
translocation event
der(4)
der(14)

Figure C4.2 – The Different IgH Translocation Models in Myeloma

A non-scale diagram of an IgH locus (14q32) and the 4p16 region involved in t(4;14) are shown on top. To limit the illustration size, the intervening region between the IgD constant region and switch gamma2 are not shown. The *LETM1* exons are indicated by shaded boxes while the *MMSET* exons are indicated by clear boxes. The sense and anti-sense strands of the IgH locus and *MMSET* are indicated by red and blue lines, respectively. Breakpoint sites are indicated by grey vertical dashed lines. A) The classic "primary switch" model is illustrated with an illegitimate CSR involving Sμ and Sγ4. B) The "secondary switch" model is illustrated in which a cell with a legitimate Sμ-Sγ4 junction undergoes a second illegitimate CSR involving Sγ4 and Sα2. C) The "CSR refinement" model is illustrated with an initial translocation event involving Sμ and a subsequent refinement of the der(14) by a CSR between Sμ and Sγ4. D) The proposed "post-switch strand invasion" model is illustrated with an initial legitimate CSR between Sμ and Sγ4 followed by a non-CSR mediated translocation involving the rearranged Sμ.


JH6
Eμ
Iμ
Sμ
Cμ
Cδ
Sγ3
Cγ3
Sγ1
Cγ1
Sα1
Cα1
Sγ2
Cγ2
Sγ4
Cγ4
Sε1
Cε1
Sα2
Cα2
Eα2
FGFR3
LETM1
1a
1
3
4
4a
5
6
MMSET
NCI-H929
der(4)**
der(14)
KMS-18
Unable to clearly define as limited sequence was published by Ronchetti et al.
der(4)*
der(14)
?
KMS-26
der(4)
der(14)
?
Edm 1174
der(4)
der(14)
?
Inversion of Sμ
Edm 1850
der(4)
der(14)
Inversion of Sμ
Edm 434
der(4)
der(14)
?

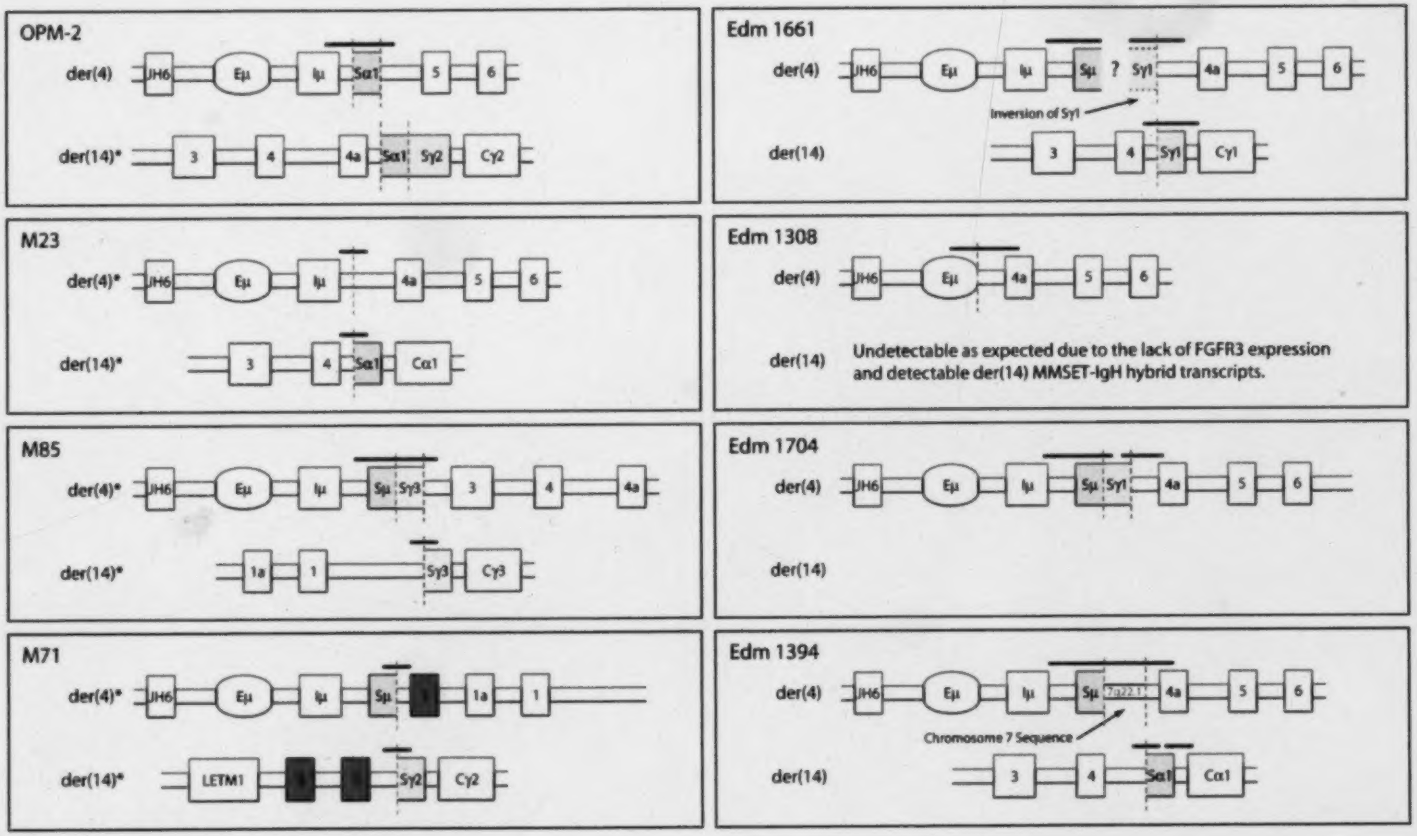

OPM-2
der(4)
JH6
Eμ
Iμ
Sα1
5
6
der(14)*
3
4
4a
Sα1
Sγ2
Cγ2
M23
der(4)*
JH6
Eμ
Iμ
4a
5
6
der(14)*
3
4
Sα1
Cα1
M85
der(4)*
JH6
Eμ
Iμ
Sμ
Sγ3
3
4
4a
der(14)*
1a
1
Sγ3
Cγ3
M71
der(4)*
JH6
Eμ
Iμ
Sμ
1a
1
der(14)*
LETM1
Sγ2
Cγ2
Edm 1661
der(4)
JH6
Eμ
Iμ
Sμ
?
Sγ1
4a
5
6
Inversion of Sγ1
der(14)
3
4
Sγ1
Cγ1
Edm 1308
der(4)
JH6
Eμ
4a
5
6
der(14)
Undetectable as expected due to the lack of FGFR3 expression and detectable der(14) MMSET-IgH hybrid transcripts.
Edm 1704
der(4)
JH6
Eμ
Iμ
Sμ
Sγ1
4a
5
6
der(14)
Edm 1394
der(4)
JH6
Eμ
Iμ
Sμ
7q22.1
4a
5
6
Chromosome 7 Sequence
der(14)
3
4
Sα1
Cα1

Figure C4.3 – Breakpoint Diagrams

Non-scale illustrations of the 14q32 and 4p16 regions involved in the translocation and diagrams of the breakpoints in each individual sample are shown. The red lines indicate the sense strand and blue lines indicate the anti-sense strand. Inversed sequence is indicated by dashed lines. The *LETM1* exons and *MMSET* exons are indicated by grey and clear boxes, respectively. Sequences generated by other groups are identified by asterisks. In one case indicated by two asterisks, NCI-H929 der(4), the breakpoint was cloned independently by ourselves and the Bergsagel group (Mayo Clinic, Scottsdale). Breakpoint junctions between two switch regions or between 14q32 and 4p16 are indicated by vertical grey dashed lines. Black bars above each illustration indicate the sequenced regions. Question marks are used to define unsequenced regions with a predicted junction between two switch regions except in the case of KMS-18 der(4) where the published sequence aligns to multiple locations on chromosome 4 and 14. All of the breakpoints cloned by our group are shown, along with selected samples from other groups for which the breakpoint of both derivatives was cloned. The analyzed breakpoint sequences and when available the associated NCBI accession numbers are as follows: M23 der(4), M23 der(14), M85 der(4), M85 der(14), M71 der(4), M71 der(14) were extracted from Fenton et al. (2003)[89]; KMS-18 der(4) was extracted from Ronchetti et al. (2001)[243]; NCI-H929 der(4)(U73662), OPM-2 der(14)(AF006657) were submitted to NCBI by the Bergsagel group (Mayo Clinic, Scottsdale); NCI-H929 der(4)(DQ090918), NCI-H929 der(14)(DQ090919), KMS-18 der(14)(DQ000667), KMS-26 der(4)(DQ000668), KMS-26 der(14)(DQ000669), OPM-2 der(4)(DQ000670), Edm 1174 der(4)(DQ090920), Edm 1174 der(14)(DQ090921), Edm 1850 der(4)(DQ090934), Edm 1850 der(14)(DQ090936), Edm 1308 der(4)(DQ090931), Edm 1704 der(4)(DQ090932), Edm 1394 der(4)(DQ090923), Edm 434 der(4)(DQ090925), Edm 434 der(14)(DQ090927), Edm 1661 der(4)(DQ090928), Edm 1661 der(14)(DQ090930), and Edm 1394 der(14)(DQ090924) were submitted to NCBI by our group.

Table C4.2 – List of Patients Fitting Each Translocation Model

	Classical	Non-Classical
	"Primary Switch"	"CSR Refinement" or "Post-Switch Strand Invasion"
Both Derivatives Cloned	Edm 1394*	Edm 434
	NCI-H929	Edm 1174
	M23	Edm 1661
	M71	Edm 1850
	JIM3†	KMS-26
	PCL-1†	OPM-2
		M85
Total	6/13 (46.2%)	7/13 (53.8%)
"Predicted" Only One Derivative Cloned	Edm 1308	Edm 1704
	KMS-18	KHM-11
	UTMC-2†	KMS-11
	T9283†	LB375
	LB1017	M65
	M57	KMS-28BM
	M62	
	M90	
	Man1	
Total	15/28 (53.6)	13/28 (46.4%)

Patients were separated into groups having either classical or non-classical translocations based on the cloned der(4) and der(14) breakpoints. The breakpoints are classified as classical translocations if the der(4) is a switch mu-chromosome 4 junction and the der(14) is a downstream switch region joined directly to chromosome 4. The breakpoints are classified as non-classical translocations if the der(4) contains a downstream switch region or the der(14) contains switch mu. The translocations detected in patients with only a single cloned derivative can also be assigned to the different translocation models. The patients classified as having non-classical translocations are definitely classified correctly as they meet the criteria outlined above. However, some of the patients classified as having classical translocations may actually have non-classical breakpoints as the other derivative, which was not cloned, may still fit the non-classical translocation model. The analyzed breakpoint sequences and, when available, their associated NCBI accession numbers are as follows: M23 der(4), M23 der(14), M62 der(4), M85 der(4), M85 der(14), M71 der(4), M71 der(14), M65 der(4), M90 der(14), M57 der(4), and Man1 der(4) were extracted from Fenton et al. (2003)[89]; LB1017 der(14) and LB375 der(14) were extracted from Richelda et al. (1997)[96]; KMS-28BM der(4) was extracted from Intini et al. (2004)[236]; KHM-11 der(4) was extracted from Sonoki et al. (2004)[247]; KMS-11 der(14)(U73663), PCL-1 der(4)(U73661), NCI-H929 der(4)(U73662), JIM3 der(14)(U73660), OPM-2 der(14)(AF006657) were submitted to NCBI by the Bergsagel group (Mayo Clinic, Scottsdale); NCI-H929 der(4)(DQ090918), NCI-H929 der(14)(DQ090919), KMS-18 der(14)(DQ000667), KMS-26 der(4)(DQ000668), KMS-26 der(14)(DQ000669), OPM-2 der(4)(DQ000670), Edm 1174 der(4)(DQ090920), Edm 1174 der(14)(DQ090921), Edm 1850 der(4) (DQ090934), Edm 1850 der(14)(DQ090936), Edm 1308 der(4) (DQ090931), Edm 1704 der(4)(DQ090932), Edm 1394 der(4)(DQ090923), Edm 434 der(4)(DQ090925), Edm 434 der(14)(DQ090927), Edm 1661 der(4)(DQ090928), Edm 1661 der(14)(DQ090930), and Edm 1394 der(14)(DQ090924) were submitted to NCBI by our group.

* - Classified as classical even though a 169 bp insertion from 7q22.1 exist on der(4) between the switch mu and *MMSET* intron 4 junction.

† - Not all sequences are publicly available, however, the relative positions of one or both derivatives were reported in Bergsagel and Kuehl (2001)[232].

C4.3.4 – Analysis of the Cloned t(4;14)(p16;q32) Breakpoints

The individual breakpoint sequences were aligned to human genome build 35.1 contigs containing the IgH locus (NG_001019) and the 4p16 region associated with the translocation (NT_037623.4) (Table C4.3). We were able to precisely define the breakpoint location on chromosome 4 and at a minimum the involved IgH switch region for all of the breakpoints that were cloned. In some instances the precise breakpoint in the associated switch region could not be determined due to the highly repetitive nature of these genomic regions. Moreover, we analyzed all but one of the publicly available t(4;14) breakpoint sequences. In general the der(4) and der(14) breakpoint junctions are well conserved with small regions of microhomology (mean, 2; range, 1-4 nt) or small non-template additions (mean, 8; range, 1-11 nt) at the junction (Table C4.3). When only patients with cloned der(4) and der(14) breakpoints are considered, we can define the sequence loss or gain from each respective chromosome. The chromosome 4 sequence is highly conserved with 7 of 12 patients having small deletions (mean, 18; range, 1-90 bp) and 5 of 12 patients having small duplications (mean, 12; range, 2-42 bp). The conservation of chromosome 14 sequence is dependent on which type of translocation event occurred; classical "primary switch" or non-classical "CSR refinement or post-switch strand invasion". In patients fitting the classical model, a large section of the IgH locus between the two switch regions is lost. However, in patients fitting the non-classical models with precisely defined breakpoints on both derivatives, the sequence is generally well conserved with 4 samples having small deletions (range 7-130 bp) and 1 sample (OPM-2) being perfectly conserved. Therefore, in the patients with non-classical translocations, the extremely small to non-existent loss of sequence from the hybrid switch junction argues against a CSR mediated mechanism.

Furthermore, by mapping the precise breakpoint locations we were able to refine the chromosome 4 breakpoint region to an approximate 65 kb region overlapping *LETM1* and *MMSET* (Table C4.3 and Figure C4.4). Surprisingly, the breakpoints are not randomly distributed through this 65 kb region; two hotspots contain 9/25 (36%) of the cloned breakpoints. The first hotspot is telomeric of *LETM1* exon 1 and contains 5 breakpoints within a 700 bp region. The second hotspot is within *MMSET* intron 4 and contains 4 breakpoints within a 250 bp region. Therefore, 36% of the cloned breakpoints map within regions representing less than 1.5% of the defined breakpoint region. A much larger set of samples is needed to clearly define these hotspots and the relative frequency of breakpoints within these sites. Interestingly, the MB4-2 breakpoint region only represents 5.6% of the current 64 468 bp breakpoint region defined by M62 and OPM-2, but 13.7% of our t(4;14) positive patients (See Appendix I) have this breakpoint.

Table C4.2 – Tabulated Breakpoint Information

Patient or Cell Line	der(4)			der(14)			Comparison of Both Derivatives	
	Last Nt Chr 14	First Nt Chr 4	Joint Feature	First Nt Chr 14	Last Nt Chr 4	Joint Feature	Chromosome 4 Sequence	Chromosome 14 Sequence
M62†	963689 Smu	353698	Perfect Conservation					
KMS-28BM†	ISTD "Sgamma1"	360121	1 nt microhomology					
KMS-11†				965327 Smu	360419	3 nt microhomology		
M71†	965457 Smu	360558	1 nt microhomology	Sgamma2	360551	1 nt microhomology	6 bp deletion	Smu and Sgamma2 breaks Switch Deletion
M90†				1112702 Salpha1	360664	1 nt non-template		
LB1017†				ISTD "Salpha1"	360802	3 nt microhomology		
Man1†	965185 Smu	364360	4 nt microhomology					
M65†	1112536 Salpha1	368783	2 nt microhomology					
PCL-1†	964282 Smu	377148	2 nt microhomology					
KHM-11†	1112189 Salpha1	379298	2 nt microhomology					

M85†	1049470 Sgamma3	397526	Perfect Conservation	1049490 Sgamma3	397522	3 nt non-template	3 bp deletion	Both breaks in Sgamma3 19 bp deletion
Edm 1174	963145 Smu	406472	Perfect Conservation	964110 Smu	406453	2 nt microhomology	18 bp deletion	Both breaks in Smu Inversion/Can't Define
Edm 1850	963003 Smu	407629	3 nt microhomology	963519 Smu	407631	2 nt microhomology	3 bp duplication	Both breaks in Smu Inversion 1 + 9 bp deletion
NCI-H929‡	964461 Smu	407964	3 nt microhomology	1113321 Salpha1	407957	1 nt microhomology	6 bp deletion	Smu and Salpha1 breaks Switch Deletion
JIM3†§	ISTD "Smu"	408606	1 nt non-template	ISTD "Sgamma2"	408600	Perfect Conservation	5 bp deletion	
KMS-18†	ISTD	ISTD		1176507 Sgamma2	408774	8 nt non-template		
M57†	962803 5' Smu	410703	7 nt non-template					
Edm 1661	?1078149? Sgamma1	410763	2 nt microhomology	1078184 Sgamma1	410761	1 nt non-template	1 bp deletion	Both breaks in Sgamma1 Potential duplication of at least 79 bp
M23†	962669 5' Smu	410905	Perfect Conservation	1113328 Salpha1	410913	2 nt microhomology	9 bp duplication	Smu and Salpha1 breaks Switch Deletion
Edm 434	1078367 Sgamma1	410931	2 nt microhomology	1078454 Sgamma1	410934	3 nt non-template	4 bp duplication	Both breaks in Sgamma1 87 bp deletion
KMS-26	963261 Smu	412955	2 nt microhomology	963391 Smu	412997	1 nt microhomology	42 bp duplication	Both breaks in Smu 130 bp deletion
Edm 1308	961491 Emu	413698	11 nt non-template					

Edm 1704	ISTD Sgamma1	414246	2 nt microhomology					
Edm 1394	962939 Smu	416555	insertion of 169 bp from 7q22.1 with 1 nt microhomologies on either side	1113158 Salpha1	416556	2 nt microhomology	2 bp duplication	Smu and Salpha1 breaks Switch Deletion
OPM-2†	1113096 Salpha1	418163	3 nt microhomology	1113097 Salpha1	418073	8 nt non-template	90 bp deletion	Both breaks in Salpha1 Perfect Conservation

The breakpoint sequences were aligned to NT_037623.4 (4p16) or NG_001019.4 (14q32) using the pair-wise BLAST algorithm (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) and the specific breakpoint location on each respective contig is indicated. Patients with defined or predicted non-classical breakpoints are highlighted in grey. The samples are ordered based on their breakpoint location on chromosome 4 from telomere to centromere. For some samples the available sequence was not sufficient to identify the involved switch region and thus we noted the proposed switch region in parentheses. The only publicly available sequence not analyzed is the LB375 der(14) sequence from Richelda et al. (1997)[96], as the published sequence does not align to the 4p16 contig nor does it align with any sequence in the NCBI database mapping to the 4p chromosome arm.

‡ - The der(4) sequence for NCI-H929 was independently cloned by our group and the Bergsagel group (Mayo Clinic, Scottsdale).

† - Sequence for one or both derivatives extracted from published material or NCBI[88,89,96,236,243].

§ - A small portion of the JIM3 der(4) sequence was kindly provided by Marta Chesi and Leif Bergsagel upon request.

ISTD - Insufficient Sequence To Determine. Due to the limited amount of published or available sequence the specific breakpoint could not be determined as the sequence aligned to multiple locations with equal identity.

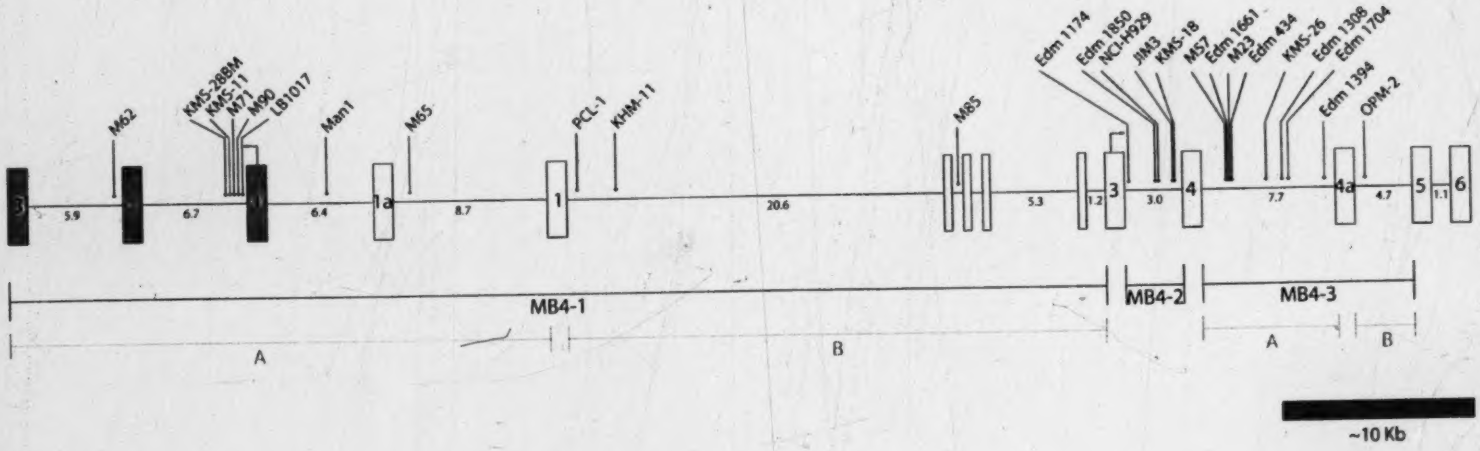

M62
KMS-28BM
KMS-11
M71
M90
LB1017
Man1
M65
PCL-1
KHM-11
M85
Edm 1174
Edm 1850
NCI-H929
JIM3
KMS-18
M57
Edm 1661
M23
Edm 434
KMS-26
Edm 1308
Edm 1704
Edm 1394
OPM-2
3
5.9
6.7
6.4
1a
8.7
1
20.6
5.3
1.2
3
3.0
4
7.7
4a
4.7
5
1.1
6
MB4-1
MB4-2
MB4-3
A
B
A
B
~10 Kb

Figure C4.4 – Characterized t(4;14) Breakpoints are not Randomly Distributed

A scale diagram of the 4p16 genomic region, based on human genome project contig NT_037623.4, with characterized t(4;14) genomic breakpoints is shown. The *LETM1* exons are shown as grey boxes and the *MMSET* exons are shown as clear boxes. The breakpoint regions identified by the der(4) IgH-MMSET hybrid transcript assays are shown by black boxed lines and the A and B sub-groups of the MB4-1 and MB4-3 classes are shown by grey boxed lines. The genomic breakpoint locations are indicated by arrows and the precise location of the breakpoint on contig NT_037623 is listed in Table C4.3. The analyzed breakpoint sequences and when available their associated NCBI accession numbers are as follows: M23 der(4), M23 der(14), M62 der(4), M85 der(4), M85 der(14), M71 der(4), M71 der(14), M65 der(4), M90 der(14), M57 der(4), and Man1 der(4) were extracted from Fenton et al. (2003)[89]; LB1017 der(14) was extracted from Richelda et al. (1997)[96]; KMS-28BM der(4) was extracted from Intini et al. (2004)[236]; KHM-11 der(4) was extracted from Sonoki et al. (2004)[247]; KMS-11 der(14)(U73663), PCL-1 der(4)(U73661), NCI-H929 der(4)(U73662), JIM3 der(14)(U73660), OPM-2 der(14)(AF006657) were submitted to NCBI by the Bergsagel group (Mayo Clinic, Scottsdale); NCI-H929 der(4)(DQ090918), NCI-H929 der(14)(DQ090919), KMS-18 der(14)(DQ000667), KMS-26 der(4)(DQ000668), KMS-26 der(14)(DQ000669), OPM-2 der(4)(DQ000670), Edm 1174 der(4)(DQ090920), Edm 1174 der(14)(DQ090921), Edm 1850 der(4)(DQ090934), Edm 1850 der(14)(DQ090936), Edm 1308 der(4)(DQ090931), Edm 1704 der(4)(DQ090932), Edm 1394 der(4)(DQ090923), Edm 434 der(4)(DQ090925), Edm 434 der(14)(DQ090927), Edm

1661 der(4)(DQ090928), Edm 1661 der(14)(DQ090930), and Edm 1394 der(14)(DQ090924) were submitted to NCBI by our group.

C4.3.5 – Description of the t(4;14) Breakpoints Cloned from the Edmonton Cohort

Edm 1174

This IgG producing myeloma patient has detectable MB4-2 bands in both der(4) hybrid transcript assays and a der(14) hybrid transcript is detect with an IgG constant region primer. As predicted by the der(4) hybrid transcript assays, the chromosome 4 breakpoint on der(4) was in *MMSET* intron 3 and joined to Sμ. Surprisingly, the der(14) did not contain a *MMSET* intron 3-switch gamma junction as predicted by the der(14) hybrid transcript assay, but was instead joined to an inverted Sμ sequence followed by a predicted inverted Sμ-Sγ3 junction. Unfortunately, bi-directional sequencing attempts did not identify the predicted Sμ-Sγ3 junction nor did they define how much Sμ sequence was inverted. The inverted Sμ sequence starts 965 bp downstream of the der(4) breakpoint so we can not determine if any chromosome 14 sequence is lost. Based on the "post-switch strand invasion" model, we predict a reciprocal translocation event occurred within the Sμ portion of a previously rearranged Sμ-Sγ3 hybrid switch region.

Edm 1850

This lambda light chain producing myeloma patient has detectable MB4-2 bands in both der(4) hybrid transcript assays and a der(14) hybrid transcript is detected with an IgA constant region primer. As predicted by the der(4) hybrid transcript assays the chromosome 4 breakpoint on der(4) is in *MMSET* intron 3 and joined to Sμ (Figure

C4.3). However, there is an inversion present within Sμ, which comprises 506 bp of downstream Sμ sequence. Surprisingly, the der(14) breakpoint did not contain the *MMSET* intron 3-switch alpha junction predicted by the der(14) hybrid transcript assay, but rather a *MMSET* intron 3-Sμ junction followed by a Sμ-Sα1 junction (Figure C4.3). Interestingly, the der(14) Sμ breakpoint is 10 nt downstream of the inverted Sμ sequence detected in the der(4), suggesting an internal Sμ inversion was associated with the translocation event Based on the "post-switch strand invasion" model we predict a reciprocal translocation event occurred within the Sμ portion of a previously rearranged Sμ-Sα1 hybrid switch region.

Edm 434

This IgG-lambda producing smoldering myeloma patient has detectable MB4-3a bands in both der(4) hybrid transcript assays and a der(14) hybrid transcript is detected with an IgG constant region primer. Surprisingly, the cloned der(4) breakpoint did not contain the Sμ-*MMSET* intron 4 junction predicted by the der(4) hybrid transcript assays, having instead a predicted Sμ-Sγ1 junction followed by a Sγ1-*MMSET* intron 4 junction (Figure C4.3). Unfortunately, our bi-directional sequencing attempts did not identify the predicted Sμ-Sγ1 junction and thus based on the location of Sγ1 in the IgH locus it is also possible that a Sμ-Sγ3-Sγ1-*MMSET* intron 4 junction exists. As predicted by the der(14) hybrid transcript assay the der(14) breakpoint consisted of a *MMSET* intron 4-Sγ1 junction (Figure C4.3). Based on the "post-switch strand invasion" model we predict a reciprocal translocation event occurred within the Sγ1 portion of a previously rearranged Sμ-Sγ1 hybrid switch region.

Edm 1661

This IgG-kappa producing myeloma patient has detectable MB4-3a bands in both der(4) hybrid transcript assays and a der(14) hybrid transcript is detected with an IgG constant region primer. Surprisingly, the cloned der(4) breakpoint did not contain the Sμ-*MMSET* intron 4 junction predicted by the der(4) hybrid transcript assays but a predicted Sμ-inverted Sγ1 junction followed by a inverted Sγ1-*MMSET* intron 4 junction (Figure C4.3). Unfortunately, our bi-directional sequencing attempts did not identify the predicted Sμ-inverted Sγ1 junction nor did it define the specific breakpoint site for the inverted Sγ1. Therefore, based on the available information and the position of Sγ1 in the IgH locus, the der(4) junction is most likely one of following possibilities: a Sμ-inverted Sγ1-*MMSET* intron 4 junction, a Sμ-Sγ1-inverted Sγ1-*MMSET* intron 4 junction, or a Sμ-Sγ3-inverted Sγ1-*MMSET* intron 4 junction. As predicted by the der(14) hybrid transcript assay the der(14) breakpoint consisted of a *MMSET* intron 4-Sγ1 junction (Figure C4.3). Based on the "post-switch strand invasion" model we predict a reciprocal translocation event occurred within the Sγ1 portion of a previously rearranged Sμ-Sγ1 hybrid switch region.

Edm 1308

This IgA-lambda producing myeloma patient is unusual, with detectable MB4-3a bands in the JH-MMSET but not in the Iμ-MMSET hybrid transcript assays. Moreover, this patient does not express detectable levels of *FGFR3* nor are der(14) hybrid transcripts detected with any IgH constant region primer. Based on these results we predicted the chromosome 14 breakpoint on der(4) occurred in a region between the JH and Iμ loci, while the der(14) is likely lost or was never formed. Exactly as predicted, we detected a der(4) breakpoint with a junction between the 3' end of the mu enhancer and *MMSET* intron 4 (Figure C4.3). The location of this breakpoint and the observed dysregulation of *MMSET* transcripts in this patient (See C2.3.5) supports the hypothesis that the oncogenic mechanism of t(4;14) is the dysregulated expression of *MMSET* transcripts by the mu enhancer. As predicted we were not able to amplify a der(14) breakpoint using any of our aforementioned strategies.

Edm 1704

This patient appears to be a unique t(4;14) case. The patient produces an IgG-kappa monoclonal protein; however, the diagnostic marrow was infiltrated with small mature CD20 positive lymphocytes representing 30% of the marrow cellularity. We detected MB4-3a bands in both der(4) hybrid transcript assays, but were unable to detect a der(14) hybrid transcript with any IgH constant region primer even though *FGFR3* transcripts are detectable in this patient. Surprisingly, the cloned der(4) breakpoint did not contain the Sμ-*MMSET* intron 4 junction predicted by the der(4) hybrid transcript assays but a Sμ-Sγ1 junction followed by an Sγ1-*MMSET* intron 4 junction (Figure C4.3). Unfortunately, even though this patient expresses *FGFR3*, we were not able to clone the der(14) breakpoint with any of our aforementioned strategies and in particular with an IgG specific reaction. Though only one derivative is cloned this patient does not fit the classic "primary switch" model and most likely fits the "post-switch strand invasion" model with a translocation event occurred within the Sγ1 portion of a previously rearranged Sμ-Sγ1 hybrid switch region.

Edm 1394

This IgG-kappa expressing myeloma patient has detectable MB4-3a bands in both der(4) hybrid transcript assays and, unexpectedly, a der(14) hybrid transcript was detected with an IgA constant region primer. Amazingly, the cloned der(4) breakpoint did not contain the predicted Sμ-*MMSET* intron 4 junction having instead an Sμ-7q22.1-*MMSET* intron 4 junction. To our knowledge a derivative chromosome breakpoint containing sequence from a tertiary chromosome has never been reported in myeloma. The inserted 7q22.1 sequence is 169 bp long and situated between *LOC401394* and *FLJ20013* on human genome build 35.1 contig NT_007933.14. As predicted by the der(14) hybrid transcript assay, the der(14) breakpoint consisted of a *MMSET* intron 4-Sα1 junction (Figure C4.3).

C4.3.6 – Description of the t(4;14) Breakpoints Cloned from Myeloma Cell Lines

KMS-26

This lambda producing cell line originated from a pleural effusion of a male myeloma patient producing an IgG-kappa monoclonal protein[248]. The cell lines has detectable MB4-3a bands in both der(4) hybrid transcript assays and a detectable der(14) hybrid transcript with an IgG constant region primer. As predicted by the der(4) hybrid transcript assays, we cloned a der(4) breakpoint with an Sμ-*MMSET* intron 4 junction (Figure C4.3). Unexpectedly, the cloned der(14) breakpoint did not contain the predicted *MMSET* intron 4-switch gamma junction, but rather an *MMSET* intron 4-Sμ junction followed by a predicted Sμ-Sγ1 junction (Figure C4.3). Unfortunately, our bi-directional sequencing attempts did not identify the predicted Sμ-Sγ1 junction, so the possibility remains that an *MMSET* intron 4-Sμ-Sγ3-Sγ1 junction exists. Based on the "post-switch strand invasion" model we predict a reciprocal translocation event occurred within the Sμ portion of a previously rearranged Sμ-Sγ1 hybrid switch region.

NCI-H929

This cell line originated from a pleural effusion of a female myeloma patient producing an IgA-kappa monoclonal protein[249]. The cell line has detectable MB4-2 bands in both der(4) hybrid transcript assays and no detectable der(14) hybrid transcript with any IgH constant region primer. The der(4) breakpoint was previously cloned and published by Chesi et al.[88]. The der(4) breakpoint that we cloned as part of our control reactions was identical to that reported by Chesi, with an Sμ-*MMSET* intron 3 junction (Figure C4.3). Since no der(14) hybrid transcript is detectable, we attempted to amplify this derivative with a primer specific to the IgA constant regions, as the patient from whom the cell line was established produced an IgA monoclonal protein. This strategy lead to the successful cloning of the der(14) breakpoint contains the predicted *MMSET* intron 3-Sα1 junction (Figure C4.3). Based on this information, it is surprising that der(14) hybrid transcripts from *MMSET* exon 1 or 3 were not detected with an IgA constant region primer.

KMS-18

This lambda producing cell line originated from the leukemic plasma cells of a male myeloma patient which had changed to a lambda light chain producer from the original IgA-lambda monoclonal protein production[250]. The cell line has detectable MB4-2 bands in both der(4) hybrid transcript assays and a detectable der(14) hybrid transcript with an IgG constant region primer. We have not been able to amplify the der(4) even though this derivative was previously reported by Ronchetti et al.[243]. However, we successfully cloned the der(14) breakpoint based on the der(14) hybrid transcript result and identified the predicted *MMSET* intron 3-Sγ2 junction (Figure C4.3).

OPM-2

This lambda producing cell line originated from the leukemic plasma cells of a female myeloma patient producing an IgG-lambda monoclonal protein[251]. The cell line has detectable MB4-3b bands in both der(4) hybrid transcript assays and no detectable der(14) hybrid transcript with any IgH constant region primer. Unexpectedly, the cloned der(4) breakpoint did not contain the predicted Sμ-*MMSET* intron 4a junction. It did have a 5' Sμ-Sα1 junction followed by a Sα1-*MMSET* intron 4a junction (Figure C4.3). We did not attempt to clone the der(14) as Chesi et al. had previously cloned and published this *MMSET* intron 4a-Sα1-Sγ2 breakpoint junction[88] (Figure C4.3). Similar to NCI-H929, the lack of detectable der(14) hybrid transcripts with an IgG constant region primer is surprising given the rearrangement. Based on the "post-switch strand invasion" model we predict a reciprocal translocation event occurred within the Sα1 portion a previously rearranged 5'Sμ-Sα1-Sγ2 hybrid switch region.

C4.4.1 – Chapter Conclusions

The primary goal of this section was to clone both derivative chromosomes from a series of t(4;14) positive patients and to determine if the translocation mechanism is likely related to CSR or an alternative mechanism. Furthermore, by combining the cloned breakpoints from our group and those published by other groups we hoped to identify recurrent breakpoint sites in chromosome 4.

We were able to clone at least one breakpoint from 7 patients and 4 cell lines. The breakpoint sequence for both derivatives was determined in 5 of the patients and 3 of the cell lines when our sequences are combined with publicly available data[88]. Therefore, breakpoint sequences for both derivative chromosomes are now available for 13 t(4;14) positive samples when our data is combined with data from other groups[88,89]. However, the sequences of both breakpoints are not publicly available for PCL-1 and JIM3. When the breakpoints from these 13 samples were analyzed we were surprised to find that approximately 50% fit the classic "primary switch" model while the other 50% fit the non-classic models. This relative distribution holds even when patients with only a single cloned derivative are included in the analysis. Therefore, it appears as if t(4;14) arises from two independent yet equally frequent events. The possibility still exists that the translocation mechanism in the patients with non-classical translocations is mediated by a B-cell specific mechanism, but this is no more likely than non B-cell specific mechanisms.

We identified several recurrent breakpoint sites by mapping the precise breakpoint locations of the 25 available t(4;14) positive samples to 4p16 sequence generated by the human genome project. This refined the current t(4;14) breakpoint region, defined by the most telomeric and centromeric breakpoints, to a 64.5 kb region encompassing *LETM1* intron 2 and *MMSET* intron 4a. Surprisingly, within this 64.5 kb region we identified two breakpoint cluster regions containing breakpoints from 36% of the cloned samples, however, the two hotspots only represent 1.5% of the defined 64.5 kb breakpoint region. Therefore, it appears as if some regions of 4p16 are preferentially targeted. However, the remaining breakpoints are distributed throughout the region so the question remains if numerous hotspots exist or if these identified hotspots are preferentially cloned. To address this issue in the future a comprehensive attempt to clone the breakpoints from a large series of t(4;14) positive patients is needed.

Discussion

D.1.1 – The Frequency of t(4;14)(p16;q32) in Myeloma and MGUS

The der(4) IgH-MMSET hybrid transcript assays provide a simple and reliable means to detect t(4;14)(p16;q32) in multiple myeloma. The hybrid transcript assays are highly specific, showing a perfect concordance with the "gold standard" FISH assay in the only comparison study[101]. Throughout the published literature there are only two t(4;14) positive samples by FISH which are negative in the der(4) hybrid transcript assays[82,91]. Alternatively, a single FISH negative patient with detectable hybrid transcripts and FGFR3 expression has been identified[118]. Furthermore, the hybrid transcript assays are very sensitive tests with reported sensitivities between 1/50-1/100 and 10^{-4}-10^{-5} positive cells in the single stage and nested assays, respectively[1,101]. The small difference in the reported sensitivities may reflect differences in the assay conditions such as the amount of template cDNA, random hexamer versus oligo dT cDNA synthesis, and the PCR enzymes, or they may reflect differences between the diluted cell lines[1,101]. Finally, these highly robust assays provided additional information regarding the relative breakpoint location on chromosome 4 and in some instances the breakpoint location on chromosome 14. Therefore, the der(4) IgH-MMSET hybrid transcript assays are an excellent set of tests for determining the t(4:14) status of patients with multiple myeloma and MGUS.

We screened 306 myeloma patients from our expanded cohort with the der(4) IgH-MMSET hybrid transcript assays and identified 44 (14.4%) t(4;14) positive patients. This frequency is consistent with previous reports from FISH and RT-PCR studies of t(4;14) indicating frequencies of 10-20% in myeloma patient populations (See Table I.5). Therefore, we believe our frequency of ~15% is a good representation of the frequency of t(4;14) within multiple myeloma patients. Initially several large studies reported incidence rates of 10-12%, however, these studies suffered from one of several fundamental flaws[65,81,102]. In some cases, the FISH assay was designed to detect the der(14) which is lost in about 25% of patients[1,26,82,112]. Alternatively, the scoring procedure was biased for reciprocal rearrangements so patients with a single derivative were misclassified. Following our initial work, Keats et al. 2003[1], and the supporting work from other groups[26,82], the design and scoring criteria of most FISH assays were corrected to remove the identified biases. Therefore, current studies are not acceptable if they do not address the identified biases.

The frequency of t(4;14) in MGUS has been and in some circles continues to be a major source of controversy. Several issues arise including differences in diagnosis, the allocation of asymptomatic/smoldering myeloma patients to MGUS or myeloma groups, differences in assay design, and the relatively small size of some cohorts. Within our expanded cohort we found 2/112 (1.8%) MGUS patients have detectable der(4) IgH-MMSET hybrid transcripts. This frequency is substantially lower than the observed frequency of 14.4% in our myeloma cohort. However, in one of the larger studies by Fonseca et al. the incidence of t(4;14) in MGUS was similar to their earlier observations in myeloma, 9.0% and 10.3%, respectively[81,108]. Though this study and one small study identified similar frequencies in MGUS and myeloma, when all of the studies reported to date are combined only 16/485 (3.3%) MGUS patients are t(4;14) positive (Table D.1). Therefore, we believe our data is reflective of a true difference in the relative frequency of t(4;14) in myeloma and MGUS. The increased prevalence of t(4;14) in myeloma compared to MGUS suggests the translocation is involved in the transition from MGUS to myeloma, or promotes a myeloma phenotype. However, several cases of t(4;14) positive MGUS have been reported in which transformation to myeloma did not occur even after several years of follow-up[101,108]. In one reported case with a confirmed t(4;14) positive MGUS and myeloma phase, the time between diagnosis of MGUS and transition to myeloma was 94 months[252]. Similarly, two of our t(4;14) positive myeloma patients were previously diagnosed with MGUS. If we make the assumption that these patients were t(4;14) positive during their MGUS phase, the transition to myeloma is not immediate as it occurred after 55 and 74 months, respectively. Thus, although t(4;14) appears to promote a myeloma phenotype and may be involved in the transformation from MGUS to myeloma, it is neither necessary for the transformation, nor is it a sign of inevitable or rapid transformation.

Table D.1 – The Incidence of t(4;14) in MGUS

Study	Assay	Patients	t(4;14) Positive
Avet-Loiseau (1999)[84]	FISH der(14) CD138$^+$	100	2 (2.0%)
Malgeri (2000)[101]	RT-PCR der(4)	16	1 (6.3%)
Sibley (2002)[119]	RT-PCR der(4)	13	2 (15.4%)
Fonseca (2002)[108]	cIg-FISH der(4&14) 47% CD138$^+$	56	5 (9.0%)
Avet-Loiseau (2002)[65]	FISH der(14) CD138$^+$	168	4 (2.4%)
Rasmussen (2003)[121]	RT-PCR der(4), qRT-PCR (FGFR3) der(14)	20	0
Keats (2005)[122]	RT-PCR der(4)	112	2 (1.8%)
Totals (All)		485	16 (3.3%)

The assay type, derivative chromosome detected, and purification procedure (if used) are listed.

One major advantage of the der(4) IgH-MMSET hybrid transcript assays over the "gold standard" FISH assays is the additional information they provide regarding the relative breakpoint locations on chromosome 4 and 14. Three different major breakpoint regions on chromosome 4 can be defined based on the size of the der(4) IgH-MMSET hybrid transcript detected. Patients with MB4-1 products are predicted to have breakpoints telomeric of *MMSET* exon 3. Importantly, *MMSET* exon 3 contains the proper translation initiation site for MMSET I, MMSET II, and MMSET III. Patients with MB4-2 and MB4-3 products are predicted to have breakpoints in *MMSET* intron 3 and *MMSET* introns 4 and 4a, respectively. Since the breakpoints in these patients are downstream of *MMSET* exon 3, the overexpressed transcripts encoding MMSET I and MMSET II produce truncated protein products. Therefore, based on the type of breakpoint we can predict which patients overexpress wild-type versus truncated MMSET protein variants. Moreover, based on the defined 64.5 kb breakpoint region on chromosome 4 and the relative size of each breakpoint region, we would expect a distribution of 80.5% MB4-1, 5.6% MB4-2, and 13.9% MB4-3 if the breakpoints are randomly distributed through this region. However, when our results are combined with those of other groups the patients are distributed as follows; 59.4% MB4-1, 18.8% MB4-2, and 21.9% MB4-3 (Table D.2). Though this is only an exploratory analysis, the breakpoints do not appear to be randomly distributed through the breakpoint region. In particular, the frequency of patients with MB4-2 breakpoints is almost 4 fold higher than expected. When this is considered in conjunction with the observation that 36% of the cloned breakpoints are located in 1.5% of the defined breakpoint region, it becomes highly probable that some limited amount of targeting is occurring. Moreover, we identified 3 patients with detectable JH-MMSET and not Iμ-MMSET hybrid transcripts. Originally, we predicted the chromosome 14 breakpoint occurred between the JH and Iμ loci of the IgH locus in these patients. Based on the assumption that the dysregulation of a protein product from the *MMSET* locus by the mu enhancer was part of the t(4;14) myelomagenic program, it was necessary to predict that the breakpoints would be between the mu enhancer and Iμ. We confirmed this prediction in patient 1308 by cloning the der(4) breakpoint with a chromosome 14 breakpoint in the 3' region of the mu enhancer. Fortunately, we have quantitative expression data from CD138 purified plasma cells confirming the overexpression of transcripts encoding MMSET I, MMSET II, and RE-IIBP. It is a reasonable assumption that the chromosome 14 breakpoints in the two remaining patients are also between the mu enhancer and Iμ.

Table D.2 – The Distribution of t(4;14)(p16;q32) Breakpoint Types

Study	t(4;14) Positive	Breakpoint Type		
		MB4-1	MB4-2	MB4-3
Malgeri (2000)[101]	12	6	4	2
Sibley (2002)[119]	9	3	4	2
Keats (2003)[1]	32	22	5	5
Santra (2003)[82]	8	4	1	3
Fabris (2005)[112]	6	3	2	1
Keats (2005)[122†]	13	10	1	2
Tajima (2005)[253]	10	6	0	4
Additional Patients See Appendix I	6	3	1	2
Totals (MM) Breakpoint Frequency	96	57 (59.4%)	18 (18.8%)	21 (21.9%)

† - Only the additional samples not present in the original cohort are noted.

D.1.2 – FGFR3 Expression and the der(14) Chromosome

The originally proposed t(4;14) target gene was *FGFR3*[88,96]. In normal plasma cells and plasma cells from t(4;14) negative myeloma patients the expression of FGFR3 is normally undetectable, but in t(4;14) positive patients, *FGFR3* is brought into close proximity with the 3' regulatory regions of the IgH locus and this results in a very high level of *FGFR3* expression. In addition to the overexpression of *FGFR3* several of the initially identified t(4;14) cell lines overexpressed *FGFR3* mutants with known activating mutations[88]. Therefore, the overexpression of *FGFR3* and in particular constitutively active *FGFR3* mutants was proposed to initiate in a cell signaling cascade promoting a neoplastic phenotype. The transforming capability of wild-type and mutant *FGFR3* variants has been tested and shown in several situations[35,129,130,174,176,177]. However, transformation is generally only observed with constitutively active *FGFR3* mutants and even in this situation transformation is dependent on a high expression level[35,174,176]. Moreover, less then 6% of t(4;14) positive patients have activating mutations of FGFR3[35,36,88,96,118,119,168-171,174].

To our initial surprise, *FGFR3* expression was only detectable in 31/44 (70.5%) t(4;14) positive patients using a qualitative RT-PCR assay[1,122]. Although our original report (Keats et al. (2003)) is usually credited with this observation it was potentially observed in two previous situations. The first observation was made in the t(4;14) positive cell line JIM3, which does not express *FGFR3*, even though *FGFR3* transcripts were apparently detectable in the original tumour[88]. In the second situation, Nakazawa et al. identified a single t(4;14) positive patient by interphase FISH lacking detectable *FGFR3* expression[99]. In both cases the der(14) was detectable by RT-PCR or FISH[91,99].

This is in stark contrast to our observations where der(14), detected by der(14) MMSET-IgH hybrid transcript assays, is not detectable in any of our *FGFR3* non-expressors. This suggests that at least in our patient cohort, that the lack of *FGFR3* expression may be due to a loss of der(14) or a transcriptional block inhibiting transcription of *FGFR3* and der(14) hybrid transcripts. The large number of t(4;14) positive patients with MB4-1 breakpoints and detectable *FGFR3* expression but without detectable der(14) hybrid transcripts is likely due to breakpoints occurring telomeric of *MMSET* exon 1. In this situation *MMSET* exon 1, which contains the 5' primer, is on der(4) not der(14), and thus we would not expect to detect der(14) hybrid transcripts. This explanation is supported by our observation that 8/11 (72.7%) cloned MB4-1 breakpoints are telomeric of *MMSET* exon 1. However, 2/4 patients with MB4-2 breakpoints and detectable *FGFR3* transcripts do not have detectable der(14) hybrid transcripts. This is surprising since *FGFR3* expressing patients with MB4-2 and MB4-3 breakpoints are expected to have detectable der(14) hybrid transcripts. Several potential explanations exist including; a looping out of sequence between *MMSET* exon 1 and 4, a switch event involving IgE, and a breakpoint within either *MMSET* exon 3 or an involved IgH constant region. However, two of the myeloma cell lines with either an MB4-2 or MB4-3 breakpoint, NCI-H929 and OPM-2, also lack detectable der(14) hybrid transcripts. In both cases, we and others have cloned the der(14) breakpoints and found no obvious reason for the lack of detectable der(14) hybrid transcripts[88].

Several groups have replicated our observations and it is now widely accepted that approximately 25% of t(4;14) positive patients are FGFR3 non-expressors (Table D.3). The first group to replicate our results used a combination of gene expression profiling, der(4) IgH-MMSET hybrid transcripts, and interphase FISH[82]. This study provides a great deal of insight into the potential molecular mechanisms underlying t(4;14) positive patients lacking FGFR3 expression. The t(4;14) positive patients without detectable FGFR3 expression can be separated into three groups based on a series of interphase FISH experiments. The first group contains 3/10 patients with apparent der(14) chromosomes defined by FGFR3-IgH fusions. The second group with 3/10 patients does not have detectable der(14) chromosomes, however, a deletion of FGFR3 appears to have occurred while the IgH region is retained. The last group contains 4/10 patients were either the der(14) is lost or it never formed and the unpaired chromosomal segments are lost as deletions of FGFR3 and IgH are detected. However, these groups may not reflect an actual deletion of *FGFR3* as the 3' end of the "FGFR3" FISH probe is actually 438 kb telomeric of *FGFR3*. Unfortunately two additional studies failed to correctly address this issue, as their FISH assays are also flawed[112,254]. In the study of Fabris et al. the t(4;14) positive patients lacking *FGFR3* expression did not have detectable FGFR3-IgH fusions. Unfortunately, though they used an "FGFR3" FISH probe flanking *FGFR3*, the IgH constant region probe does not flank the 3' end of the IgH constant region. Moreover, the number of IgH or FGFR3 signals was not reported, so these patients can not be assigned to the groups identified by Santra et al[82]. In a recent study by Chang et al. all of their t(4;14) positive patients lacking *FGFR3* expression contained FGFR3-IgH fusions[80]. However, this study did not use an assay capable of identifying the der(4) and may be under representing the incidence of t(4;14) positive patients without *FGFR3* expression. Furthermore, similar to Fabris et al. the IgH constant region probe does not flank the 3' end of the IgH constant region. Therefore, a number of the t(4;14) negative patients in this study are likely false negatives due to the poorly designed assays designed to detect the genomic der(14) or a protein product originating from der(14).

Table D.3 – FGFR3 Expression in t(4;14) Positive Samples

Study	t(4;14) Positive	FGFR3 Expressors
Nakazawa (2000) [99]	7	6
Sibley (2002) [119]	7	7
Keats (2003 & 2005) [1,122]	44	31
Santra (2003) [82]	32	22
Rasmussen (2004) [121]	3	2
Fabris (2005) [112]	6	4
Tajima (2005) [253]	9	8
Chang (in press) [254]	16	12
Totals	**124**	**92 (74.2%)**

This table lists all t(4;14) positive patients identified by either FISH, RT-PCR, or gene expression profiling for which the expression of FGFR3 was determined by RT-PCR, qRT-PCR, gene expression profiling, or immunohistochemistry.

Though we and others have identified a series of t(4;14) positive patients with non-reciprocal translocations this phenomena is not limited to t(4;14). In a comprehensive study using a series of well designed FISH assays, Fonseca et al. identified a similar rate, approximately 30%, of non-reciprocal translocations in t(4;14), t(11;14), and t(14;16)[26]. Unfortunately, this study was not designed to determine which derivative chromosome is lost; it does however, exquisitely show that non-reciprocal IgH translocations are common in myeloma. The mechanism generating non-reciprocal IgH translocations is not known but a study using metaphase FISH provides some insight into two potential mechanisms[63]. The first, and most obvious is that der(14) is lost or never created. The second involves a complex rearrangement involving 3 loci resulting in a der(4)t(4;14), a der(14)t(?;14), and potentially a der(?)t(4;?) containing *FGFR3*. In this latter situation FGFR3 may persist but no der(14)t(4;14) would be detected. This highlights the absolute need for well designed and rationalized FISH assays for the detection of IgH translocations in myeloma.

D.1.3 – The Clinical Significance of t(4;14)(p16;q32) in Multiple Myeloma

One of the principal aims of this study was to determine the clinical significance of t(4;14)(p16;q32) in multiple myeloma. We investigated the clinical significance of t(4;14) in a retrospective cohort of 208 myeloma patients collected over an 8 year period between 1994 and 2002. Within this cohort of patients the translocation, as detected by der(4) IgH-MMSET hybrid transcript assays, is associated with a decrease in overall survival (P=0.037) and a decreased response to front-line chemotherapy therapy (P=0.05). The median survival of patients with t(4;14) is 709 days which is significantly reduced compared to the 1338 days observed in t(4;14) negative cases. The association between t(4;14) and a decrease in overall survival was first reported at the American Society for Hematology annual meeting in December 2001 by our group and groups from the Mayo Clinic and IFM. Subsequently, all three groups published the observation within a 10 month period between September 2002 and June 2003[1,78,83]. In all three studies the treatments were different, with the Mayo clinic and IFM studies analyzing cohorts of patients treated uniformly with conventional chemotherapy or autologous transplant following HDT while our cohort was treated with both modalities. Furthermore, the poor prognosis of t(4;14) positive patients treated with autologous transplant was recently verified by a fourth independent group[80]. Therefore, t(4;14) predicts for a poor overall prognosis in patients treated with any of the standard front-line therapies for multiple myeloma.

Though the association between t(4;14) and reduced overall survival is now well accepted, the mechanism behind this aggressive form of myeloma is not known. Since the originally proposed t(4;14) target gene, *FGFR3*, was not expressed in all of our t(4;14) positive patients, we performed a survival analysis which accounted for *FGFR3* expression status. Surprisingly, both t(4;14) positive patients with or without *FGFR3* expression had reduced median survivals compared to t(4;14) negative patients; 813 versus 692 versus 1338 days, respectively. Unexpectedly, in this initial exploratory analysis it was the t(4;14) positive patients lacking *FGFR3* expression which remained statistically significant (P=0.016). Though no significant difference in survival existed between t(4;14) positive patients with or without *FGFR3* expression in our original cohort, the difference in median survivals suggested a possible trend. Therefore, we expanded our analysis to included t(4;14) positive patients identified in the expanded cohort. Similarly, this secondary analysis did not identify a difference in overall outcome between t(4;14) positive/FGFR3 positive and t(4;14) positive/FGFR3 negative patients, median survival 813 and 709 days, respectively[122]. Therefore, the poor outcome associated with t(4;14) is independent of *FGFR3* expression, suggesting *FGFR3* is not a clinically relevant t(4;14) target gene. Since the expression of *FGFR3* did not influence the clinical outcome of t(4;14) positive myeloma patients we investigated the clinical impact of the second proposed t(4;14) target gene, *MMSET* [91]. This gene is expressed in all t(4;14) positive samples, however, the expressed transcripts do not encode full length protein variants in all patients (MB4-1 versus MB4-2/MB4-3 patients). Therefore, we performed an exploratory analysis on the t(4;14) positive patients from our expanded cohort, to determine if the ability to produce full length MMSET protein variants influence the survival of t(4;14) positive patients, and no significant difference was observed[122]. Therefore, neither the expression of FGFR3 nor the expression of wild-type or truncated forms of MMSET influence the poor prognosis associated with this translocation. This suggests that some as yet undefined and universal feature of t(4;14) is responsible for this aggressive from of multiple myeloma.

Given our observations even the limited success of FGFR3 inhibitors in pre-clinical models is surprising[36,255-260]. Interestingly, in our analysis of sequential bone marrow samples we have not observed a single t(4;14) positive patient lose *FGFR3* expression, suggesting, it may still be a pertinent factor in t(4;14) myelomagenesis even though it does not impact clinical outcome. However, these aforementioned studies were almost exclusively performed in myeloma cell lines and thus the results may not extrapolate to patients. Moreover, no cell line has been derived from a confirmed FGFR3 non-expressing patient, suggesting this feature maybe essential for the development of a t(4;14) cell line but may not be essential for the survival of t(4;14) positive cells in patients. Further work is required to justify the use of FGFR3 inhibitors in t(4;14) positive patients.

Although t(4;14) is a widely accepted poor prognostic marker several discrepancies exist. When *FGFR3* expression is used as a marker for t(4;14), neither our group nor Rasmussen et al. observed a significant difference in survival between *FGFR3* expressors and non-expressors[1,120]. Most likely this discrepancy reflects the impact of the t(4;14) positive/FGFR3 negative patients on the outcome of the *FGFR3* negative group. Interestingly, a recent report by Chang et al. using immunohistochemistry to detect FGFR3 expression identified a significant difference in survival between FGFR3 expressors and non-expressors[254]. Though this result disagrees with the aforementioned studies it may simply reflect differences in the cohorts. Furthermore, it lends support to the previous studies that identified a clinical impact for t(4;14) detected by der(14) specific interphase FISH assays[80,83].

In our original cohort of 208 myeloma patients we did not find a correlation between t(4;14) and elevated beta-2-microglogulin levels or an IgA monoclonal protein as initially suggested by Avet-Loiseau et al[1,65]. The association between t(4;14) and elevated beta-2-microglobulin levels has not been observed in any cohort other than those reported by the IFM collaborative group[1,26,65,80,83,120]. The association between t(4;14) and an IgA monoclonal protein is controversial, however, in retrospect this may relate to the detection strategy and the associated number of false negatives in der(14) specific interphase FISH assays. No correlation was observed between t(4;14) and IgA in our cohort or the cohort reported by Fonseca et al[1,26]. In both cases, the t(4;14) screening assays could detect der(4), which is detected more consistently than der(14)[82,112]. Conversely, the studies showing a correlation between t(4;14) and IgA are flawed and likely have a number of false negative t(4;14) results, as the interphase FISH assays used in these studies detect exclusively the der(14)[65,80,83,254]. Therefore, no correlation between t(4;14) and a clinical pathological feature has been reported by independent groups using a highly accurate assay. One potentially valid correlation is an observed association between t(4;14) and increased serum interleukin-6 receptor levels, however, this observation requires independent verification[26].

Since data on other chromosomal abnormalities are not available for this cohort, we cannot comment on the well established correlation between t(4;14) and chromosome 13 abnormalities. Never the less, we expect our cohort is similar to those reported by other groups showing a strong correlation[26,65,81,83]. However, the independent prognostic significance of t(4;14) in a multivariable analysis adjusting for chromosome 13 abnormalities has been noted in one study and suggested in another[26,83]. Therefore, even though t(4;14) correlates with chromosome 13 abnormalities, it is the translocation that appears to confer an aggressive form of the disease.

One of the strengths of our study is that it is population-based, and the cohort is therefore representative of all patients who presented to the myeloma clinics in our region. Thus the results can be generalized. However, our clinical data was collected retrospectively, and therefore, subject to some limitations. In particular, the availability of data on some baseline clinical features and response to therapy was somewhat limited. The prognostic impact of t(4;14) in specific patient subgroups, and in particular the impact of t(4;14) on response to therapy, needs to be confirmed in large prospective studies. We are also limited as t(4;14) status was not determined at the time of diagnosis for all patients, and thus we have assumed, based on the available evidence, that t(4;14) status does not change from the time of diagnosis. It is reassuring to note that the prognostic significance of t(4;14) remains even when only those patients for whom t(4;14) status is known at diagnosis are considered.

D.1.4 – The Overexpression of MMSET Transcripts in t(4;14)(p16;q32) Myeloma

The region of chromosome 4 involved in t(4;14) is part of a conserved gene cluster originating from three *en bloc* duplications of a common orthologous region. These events generated four paralogous regions on 8p11, 10q26, 4p16, and 5q35 with each cluster being defined by *FGFR1*, *FGFR2*, *FGFR3*, and *FGFR4* respectively[126,128]. This gene cluster includes the TACC gene family members, which are numbered to match the associated FGFR paralog with *TACC1*, *TACC2*, and *TACC3* located at 8p11, 10q26, and 4p16, respectively. The other major member of the gene cluster is the NSD gene family with *NSD1*, *WHSC1/MMSET/NSD2*, and WHSC1L1/NSD3 located at 5q35, 4p16, and 8p11, respectively. Based on their proximity to known t(4;14) genomic breakpoints *FGFR3*, *MMSET*, and *TACC3* were proposed as potential t(4;14) target genes by other groups[88,90-92]. Moreover, based on our initial work, we suggested *LETM1* and *WHSC1* as potential t(4;14) target genes due to their proximity to known t(4;14) genomic breakpoints[1].

All of the cloned t(4;14) genomic breakpoints are centromeric of *FGFR3* and as a result the *TACC3* and *FGFR3* are brought into close proximity with the 3' regulatory regions of the IgH locus on der(14). Alternatively, *MMSET* and *WHSC2* are brought into proximity of the mu enhancer on der(4). However, depending on the breakpoint location parts of the 5' UTR or even translated regions of *MMSET* are separated onto the different derivatives. Finally, *LETM1* will be brought into proximity of the 3' regulatory regions of the IgH locus. Still, many of the breakpoints occur within this gene and others may occur in 5' regulatory regions. Therefore, *LETM1* may be overexpressed due to proximity with IgH enhancer elements or underexpressed due to the breakpoint locations.

Since t(4;14) predicts for a poor overall outcome irrespective of associated factors we hypothesized that true t(4;14) target genes would be uniformly overexpressed or underexpressed in all t(4;14) positive patients.

We used qRT-PCR with taqman probes to identify uniformly dysregulated transcripts originating from 4p16 in all t(4;14) positive patients. Although all of the analyzed transcripts were dysregulated in some samples, the only universally dysregulated transcripts in all t(4;14) positive samples originate from the MMSET locus. This includes the *MMSET* splice variants encoding MMSET I and MMSET II along with the transcript encoding RE-IIBP, which originates from an alternative transcription event. The only *MMSET* transcript which is not universally dysregulated encodes MMSET III. This transcript is interrupted by breakpoints within *MMSET* intron 3 or 4 which are observed in patients with MB4-2 and MB4-3a breakpoints. Interestingly the overexpressed RE-IIBP transcripts detected in t(4;14) positive patients appear to be de novo transcription events originating from the RE-IIBP promoter and not IgH hybrid transcripts.

The dysregulation of potential t(4;14) target genes has been investigated by several groups and the overexpression of *MMSET* was observed in all t(4;14) positive patients in the majority of the studies[82,112,117,122]. The only exception was the study by Stewart et al., but the accuracy of these results are open to question given the observations of other groups discussed earlier[100]. Moreover, the co-authors of this study previously reported spiked expression of *MMSET* in all of their t(4;14) positive patients using gene expression profiling[82]. Unfortunately, it was not mentioned if any of analyzed patients from these two studies overlapped[82,100]. Regardless of this latter study, the overexpression of *MMSET* in t(4;14) positive patients is well accepted. Surprisingly, the major issue is becoming the expression of *MMSET* in t(4;14) negative cases. Several studies identified increased expression levels of *MMSET* in t(4;14) negative cases and in some instances the expression level approaches the levels observed in t(4;14) positive samples[100,112,117]. Similarly, the expression level of *MMSET* in one of our t(4;14) negative patients is comparable to the lowest level observed in our t(4;14) positive patients. Interestingly, Fabris et al. recently correlated this increase in *MMSET* expression with extramedullary disease[112]. Moreover, the t(4;14) negative patient with the highest expression level of *MMSET*, the genomic region encompassing *FGFR3* and *MMSET* was on an unidentified chromosome[112]. This suggests that *MMSET* may be dysregulated by mechanisms unrelated to immunoglobulin regulatory regions. Though *MMSET* is the most promising t(4;14) target gene, Stewart et al. observed a two fold increase in the expression of *TACC3* in their t(4;14) positive patients[100]. However, they suggested this limited overexpression was unlikely to play a major role in t(4;14) mediated myelomagenesis. Similarly, we detected an approximate 2 fold difference (REL 1.17 versus 3.90), however, this difference in mean values was largely due to one outlier and was not statistically significant.

Therefore, based on our qRT-PCR analysis and the gene expression profiling results from other groups *MMSET* is likely the true t(4;14) target gene. However, three different transcripts encoding different protein variants are overexpressed in all t(4;14) positive patients. Interestingly, both MMSET paralogs, WHSC1L and NSD1, are involved in NUP98 translocations in acute myeloid leukemia[180,181,183,184]. This suggests that the NSD protein family as a whole can influence, and likely promote, the malignant process by as yet undefined mechanisms.

D.1.5 – The MMSET Protein Variants

Overexpression of *MMSET* transcripts should result in increased levels of the encoded protein products and potentially downstream myelomagenic effects. However, the dysregulated *MMSET* transcripts do not encode identical protein products in all patients. In our expanded cohort 30% of the t(4;14) positive patients have genomic breakpoints downstream of *MMSET* exon 3, separating the first, or first and second, translated exons from the remaining translated exons. In patients with MB4-2 and MB4-3 breakpoints, truncated MMSET protein variants may be produced from alternative translation initiation sites identified in *MMSET* exons 4 and 6[91]. Immunoblot analysis of HeLa cells transiently transfected with C-terminally tagged GFP constructs representing the overexpressed MMSET I and MMSET II encoding transcripts in MB4-2 and MB4-3 patients identified bands corresponding to translation products from the predicted alternative translation initiation sites. Therefore, the predicted alternative translation initiation sites are used by MB4-2 and MB4-3 patients to produce truncated MMSET protein variants. Identical observations were made in immunoblot experiments on t(4;14) positive cell lines using antibodies specific to MMSET (Leif Bergsagel, Pers. Comm.).

Localization of MMSET I and II with GFP constructs confirmed the predicted nuclear localization. The constructs with N-terminal GFP tags were excluded from the nucleoli while the constructs with C-terminal tags were largely nucleoplasmic with weak staining of the nucleoli. This discrepancy was easily explained once the localization of the MB4-2 and MB4-3 variants was identified as nuclear and enriched in nucleoli. We believe the wild-type/MB4-1 MMSET variants generally use the proper translation initiation site in *MMSET* exon 3, but on rare occasions use the alternative translation initiation sites in *MMSET* exons 4 and 6 to produce protein products that are enriched in nucleoli. In support of this assumption, anti-MMSET immunoblotting of protein extracts from myeloma cell lines with MB4-1 breakpoints identifies a predominant band corresponding to the predicted full length protein variants (Leif Bergsagel, Pers. Comm.). Therefore, an essential domain required for the localization of MMSET protein variants is located in the N-terminus encoded by *MMSET* exons 3 and 4, which is lost in the MB4-2 and MB4-3 breakpoint variants.

The majority of the N-terminal protein segment lost in MB4-2 protein variants is encoded by a naturally occurring *MMSET* splice variant. This transcript is produced by an alternative splicing event between *MMSET* exons 4 and 4a. Due to an in-frame stop codon in *MMSET* exon 4a the encoded polypeptide, MMSET III, almost perfectly represents the protein segment lost in MB4-2 variants. Confirming the presence of a regulatory domain in the N-terminus of MMSET variants, the MMSET III variant localized to the nucleus and was excluded from nucleoli. Furthermore, this protein segment is a robust mediator of the nuclear/non-nucleolar phenotype observed in wild-type MMSET variants. This was confirmed using a fusion construct of MMSET III and B23, a nucleolar protein, resulting in a mixed population of transiently transfected cells. The majority of the cells had a wild-type MMSET phenotype of nucleolar exclusion. Therefore, the N-terminus of MMSET is capable of over-riding the nucleolar localization signal present in B23.

The only characterized domain within the N-terminus of wild-type MMSET variants is the PWWP domain. The PWWP domain is a conserved domain of 50-80 amino acids identified in a variety of nuclear proteins[192,208]. This domain was recently shown to bind DNA in vitro and chromatin in vivo[205-207]. However, if the PWWP domain is the essential domain, the question remains how the limited portion of the domain retained in MMSET III is capable of mediating the localization when additional C-terminal residues are required for the proper folding of the domain[208]. Using a variety of online prediction programs, we were unable to identify a characterized protein domain N-terminal of the PWWP domain. By comparing the N-terminus of MMSET to the nearest paralog, WHSC1L1, and the evolutionarily conserved orthologs we identified several regions of highly conserved sequence. One region, amino acids 116-146 of MMSET, is highly conserved in the majority of the analyzed sequences. However, when all of the MMSET orthologs are analyzed, three highly conserved motifs are identified. Potentially these regions represent novel domains or modification and interaction motifs. However, we did not identify a known motif in any of the highly conserved regions. In the future it will be necessary to determine if one of these regions mediates the localization pattern and if so, to identify the mechanism.

The localization differences suggested a loss-of-function in the MB4-2 and MB4-3 MMSET protein variants. To determine if the localization differences reflected a functional difference, we characterized the intracellular association kinetics of each variant using FRAP. As expected, the larger proteins had the slowest recovery kinetics. Interestingly, the recovery of MMSET II is very slow and may be similar to some histone H1 variants, although direct comparisons using identical conditions have not been performed[261]. The slow recovery kinetic and the high degree of co-localization with DNA in live cells, suggests MMSET II is interacting with a chromatin component with a very high affinity. The affinity of this interaction may explain the difficulties many groups have experienced in obtaining strong bands for immunoblot experiments, as detergents used in the protein extraction step may not solubilize this protein/chromatin complex sufficiently. Moreover, if the expression level of a slow moving protein is tightly regulated, a substantial loss-of-function might be expected in haplo-insufficient cells. Consistent with this hypothesis, haplo-insufficiency of the *MMSET* paralog *NSD1* is associated with Sotos syndrome and haplo-insufficiency of *WHSC1/MMSET* is likely a causative factor in Wolf-Hirschhorn syndrome[90,262]. The recovery kinetics of the truncated MMSET I and MMSET II variants encoded by MB4-2 and MB4-3 transcripts are substantially reduced compared to the wild-type/MB4-1 MMSET variants. Therefore, not only does the N-terminus mediate the localization of wild-type MMSET variants, but it also influences the mobility of the proteins. Furthermore, the contribution of the N- and C-terminus to the mobility of MMSET II must be synergistic, as MMSET III and MB4-2 II would have an aggregate recovery comparable to MMSET II if the mobilities were additive. Therefore, if overexpression of MMSET I or MMSET II variants is the myelomagenic event in t(4:14) myeloma, it is unlikely to be related to the normal function of MMSET because the protein variants are not functionally equivalent in all t(4;14) positive patients.

Since all MMSET protein variants are located in the nucleus, the predominant localization of RE-IIBP to cytoplasmic foci with a minor pool in nucleoli was surprising. The localization pattern was consistent in transient transfections with both N- and C-terminal GFP tags and in stable transfectants with the N-terminal GFP tag. Interestingly, the localization of MB4-2 and MB4-3 MMSET I variants and RE-IIBP to nucleoli suggests there are two independent nucleolar localization mechanisms in the N- and C-terminal regions of MMSET, respectively.

RE-IIBP was initially identified as binding response element II of the IL-5 promoter *in vitro* and regulating IL-5 transcription *in vivo*[188]. Our localization data would suggest this is a minor activity since very little RE-IIBP localized to the nucleoplasm where this function is predicted to occur. Moreover, the PWWP domain is the only domain in RE-IIBP with a suggested DNA binding capability but, at least in the case of Dnmt3b, this interaction is non-specific[207]. Thus the function of RE-IIBP remains elusive and requires further study. This is particularly true given our observation that RE-IIBP is the only overexpressed protein with uncompromised function that is present in all t(4;14) positive patients. Finally, the localization of the only universally and functionally equivalent protein product to nucleoli is interesting given the observation that t(4;14) correlates with an immature or intermediate plasma cell morphology in which a large portion of identified plasma cells have nucleoli[85].

D.2.1 – Model of t(4;14)(p16;q32) Mediated Myelomagenesis

The series of events eventually culminating in multiple myeloma continues to be elusive. However, it is slowly becoming apparent that myeloma is a global term defining a series of genetically distinct diseases resulting in a similar clinical pathological phenotype. Defining these genetically distinct diseases will require substantial effort from the myeloma research community, but several fundamental features are becoming evident based on cytogenetics and gene expression profiling. A variety of cytogenetic features can be used to differentiate patients into different subgroups. Using standard metaphase cytogenetics or spectral karyotyping, patients are separated into hyperdiploid or non-hyperdiploid categories; hypodiploid, pseudodiploid, and near-tetraploid[43,44,46,63,68,97,98,263,264]. These global cytogenetic categories have prognostic impact as the overall survival of patients in the hypodiploid subgroup is significantly reduced[43,44]. Alternatively, the recurrent IgH translocations can be used to identify high risk t(4;14)/t(14;16)] and low risk t(11;14) patient subgroups[1,26,65,80,83]. Interestingly, the incidence of IgH translocations is much higher in the non-hyperdiploid category[63,78,114]. Finally, chromosome 13 deletions are common in myeloma and identify a poor prognosis group[26,43,44,47,48,50,53,65,83,84,265-268]. Similarly, global gene expression profiling of myeloma plasma cells identifies several subgroups with different survival characteristics[115,269]. However, many of the recently identified subgroups correlate with known cytogenetic features[269]. Therefore, further work is required to define each subgroup and the sequence of myelomagenic events.

Although t(4;14) was only identified 8 years ago, our understanding of this translocation is extensive. The translocation is identified in approximately 15% of myeloma cases making it the second most common IgH translocation in multiple myeloma[1,26,65,80,82,83,101,122]. Unfortunately, t(4;14) identifies a subgroup of patients with an extremely poor prognosis irrespective of the treatment strategy[1,26,80,83]. However, t(4;14) is intimately linked to two additional cytogenetic markers of poor prognosis; chromosome 13 deletions and hypodiploidy[26,63,65,81,83]. Though no conclusive evidence exists, among these abnormalities t(4;14) is generally believed to be the most significant contributor to poor prognosis. First, chromosome 13 deletions and hypodiploidy are not found in all t(4;14) positive patients nor do all patients with either abnormality have t(4;14)[63,78]. Second, chromosome 13 deletions and hypodiploidy may be intermittently detected at diagnosis and relapse while t(4;14) is universally detected at either point[1,53,122]. Third, patients with chromosome 13 deletions but not t(4;14) represent an intermediate prognostic group, suggesting t(4;14) confers additional poor prognostic features beyond those conferred by chromosome 13 deletions[26,83]. Therefore, t(4;14) is likely the most significant cytogenetic indicator of poor prognosis. The mechanism behind the poor prognosis conferred by t(4;14) is not known but we are slowly elucidating the biological features of this translocation.

At the genetic level t(4;14) and other IgH translocations in multiple myeloma were thought to be reciprocal translocations caused by illegitimate class switch recombination events[88,232]. However, initial work from our group and subsequent follow-up work by other groups has conclusively shown t(4;14) is not a reciprocal translocation event in all patients[1,26,63,82,112]. The non-reciprocal t(4;14) events are associated with the loss of der(14)t(4;14) and as such the initially proposed t(4;14) target gene, *FGFR3*, is not overexpressed[1,63,82,112]. The loss of der(14)t(4;14) is likely associated with two different events. In the first situation the der(14) is not detectable nor is the associated 4p telomere or chromosome 14. In this situation it is not known if the der(14) was formed and subsequently lost or if the der(14) never formed and the unligated chromosomal segments were lost independently. In the second situation der(14) is not detectable due to a complex rearrangement involving at least three genetic loci. In this situation a der(4)t(4;14) and der(14)t(?;14) are detected[63]. Based on the data presented by Santra et al. it is unlikely that a der(?)t(?;4) derivative formed, since some *FGFR3* non-expressors without FGFR3-IgH fusions have FGFR3 deletions but lack IgH deletions[82]. In the future it will be necessary to define how these different events are linked to the translocation models outlined in chapter 4. We predict the classical "primary switch" translocations will always form a der(14) which may be lost in some patients. Conversely, those patients with non-classical "CSR refinement or post-switch strand invasion" translocations will often have complex rearrangements or unligated chromosomal segments that are subsequently lost.

At the gene expression level several genes flanking the known genomic breakpoints on 4p16 are dysregulated in t(4;14) positive patients. However, *MMSET* is the only universally dysregulated gene from this region in all t(4;14) positive patients. Importantly, this fits with the current observations showing der(4)t(4;14) is always detectable and thus suggesting the true t(4;14) target gene exists on this derivative. At least three different transcripts are overexpressed in all t(4;14) positive patients. The overexpressed transcripts encode MMSET I, MMSET II, and RE-IIBP. However, the transcripts encoding MMSET I and MMSET II encode truncated proteins in 30-40% of t(4;14) positive patients. Although dysregulation of transcription from the MMSET locus is likely a significant event in t(4;14) myelomagenesis, the dysregulation of numerous genes characterizes t(4;14) at the global gene expression level. Two studies have identified a large number of dysregulated genes in t(4;14) myeloma[117,270]. Interestingly, other than *MMSET* the only overexpressed gene identified in both studies was the Cold Shock Domain Protein A (*CSDA*). A select group of the most significantly overexpressed genes in each individual study includes; *KLF4*, *NGFRAP1*, *MAL*, *CD99*, *IL6R*, *AREG*, *SPN*, *RASGRP1*, *RBMS1*, and *ZFP36L1*. In the future it will be necessary to confirm these observations, but at a minimum *CSDA* appears to be an important gene in t(4;14) myeloma.

At the protein level the myelomagenic mechanism of t(4;14) will require additional investigation. However, based on our observations we are proposing two possible models. In the first model the myelomagenic mechanism is mediated by the overexpression of a wild-type protein product. The only overexpressed protein fitting this model is RE-IIBP, since this protein is predicted to be functionally equivalent in all t(4;14) positive patients. Therefore, the overexpression of RE-IIBP and resulting increase in function would either mediate or select a neoplastic phenotype. In the second model the overexpressed protein would act through indirect mechanisms. In this situation all three protein variants; MMSET I, MMSET II, RE-IIBP and even the truncated MMSET I and MMSET II variants would fit the model. Therefore, the overexpression of one or all of the protein variants regardless of truncation would interfere with the normal function of MMSET or alternatively one of its important protein/chromatin partners. Future studies are required to determine which protein product or products are essential for t(4;14) mediated myelomagenesis. Furthermore, by characterizing the important protein variants and the mechanism by which they mediate myelomagenesis, it should be possible to design novel therapeutics to treat patients with this highly aggressive form of multiple myeloma.

Therefore, based on the information outlined above and other available pieces of information, I propose two slightly divergent models of t(4;14) mediated myelomagenesis depending the type of translocation event (Figure D.1). Importantly, these models are simplified predictions and several of the described features may be shared in patients with either type of translocation event. The models are built on several general observations drawn from human cancers in general and specifically multiple myeloma. The first major assumption is that t(4;14) is an initiating event. Although this is not proven it is a generally held beliefed that the IgH translocations are initiating events, since they are found in the large majority of the plasma cells within individual patients and they are observed in the pre-malignant condition MGUS. The second major assumption is that t(4;14) but also myeloma in general follows the "multiple hit" hypothesis. This assumption is based on two independent observations. First, multiple myeloma is a disease of the elderly with an average age at diagnosis of approximately 65 years, suggesting the individual "hits" require a substantial amount of time to accumulate. This is supported by the slow progression from MGUS to myeloma even after potentially substantial events such as t(4;14) and chromosome 13 deletions. The third major assumption is that antigen dependence is a common feature in the B-cell clones before and after the t(4;14) event occurred. If these B-cell clones were not antigen dependent at the time of the translocation event, we would expect the translocation to occur at an equal frequency on the functional or non-functionally rearranged IgH locus, however, it occurs nearly exclusively on the non-functional allele. The fourth major assumption is that either or both t(4;14) and the resulting overexpression of MMSET protein variants would activate the apoptotic program (ie. not tolerated) in normal B-cells. Thus, both models assume previous mutations exist in the progenitor B-cells or hematopoietic stem cells which allowed cells undergoing the t(4;14) event to survive, and allowed the associated overexpression of MMSET protein variants to be tolerated.

Therefore in patients with classical "primary switch" translocations, the model predicts the existence of pre-neoplastic features in the progenitors of any B-cell clone that remains viable after the initial t(4;14) event. Since, the accumulation of these mutations or predisposing features is predicted to be slow; we predict they most likely occur in a long lived subset of cells. These events could be accumulating in a subset of haematopoietic stem cells and/or memory B-cells since both are long lived cell subsets associated with B-cell development. Though both subsets could facilitate the slow accumulation of mutations, some of the mutations may occur during the development of B-cells during the various periods of genetic instability (ie. VDJ recombination, SHM, and CSR). Therefore, classical "primary switch" translocations may occur in a mature B-cell after SHM or in an IgM memory B-cell, either of which contain sufficient pre-neoplastic features to survive the t(4;14) event. Importantly, because it would otherwise occur on the functionally rearranged IgH locus in 50% of patients, the B-cell must remain dependent on antigen even after the translocation event. Subsequently, a series of events leads to antigen independence and a transition to MGUS or directly to overt myeloma. Though most of these events are not defined, the selection of a hypodiploid karyotype and an associated chromosome 13 deletion are likely required for the subsequent transition to either MGUS or myeloma.

Alternatively, in patients with non-classical "post-switch strand invasion" translocations the model predicts the existence of pre-neoplastic events in memory B-cells which have undergone affinity maturation and a legitimate bi-allelic class switch recombination. The model assumes the pre-neoplastic events occur in the long lived memory B-cell compartment, however, some or all may originate from a haematopoietic stem cell pool or may occur during B-cell development. Similar to patients with classical translocation events these memory B-cells must be dependent on antigen even after the translocation occurs. Subsequent events eventually lead to antigen independence and a transition to either MGUS or myeloma.

Ultimately the primary difference between the classical and non-classical models is the timing of the translocation event and class switch recombination. In the classical model the translocation occurs during CSR and is predicted to be mediated by the CSR machinery. In the non-classical "post-switch strand invasion" model the translocation occurs after CSR and the involvement of CSR machinery in the translocation event is unlikely. Finally, we did not propose a model for the non-classical "CSR refinement" model as the need for CSR mediated refinement on both derivatives is difficult to justify. However, this model was functionally validated by Kovalchuk et al. and in the future it will be necessary to determine if clonal cells exist without CSR refined translocations.

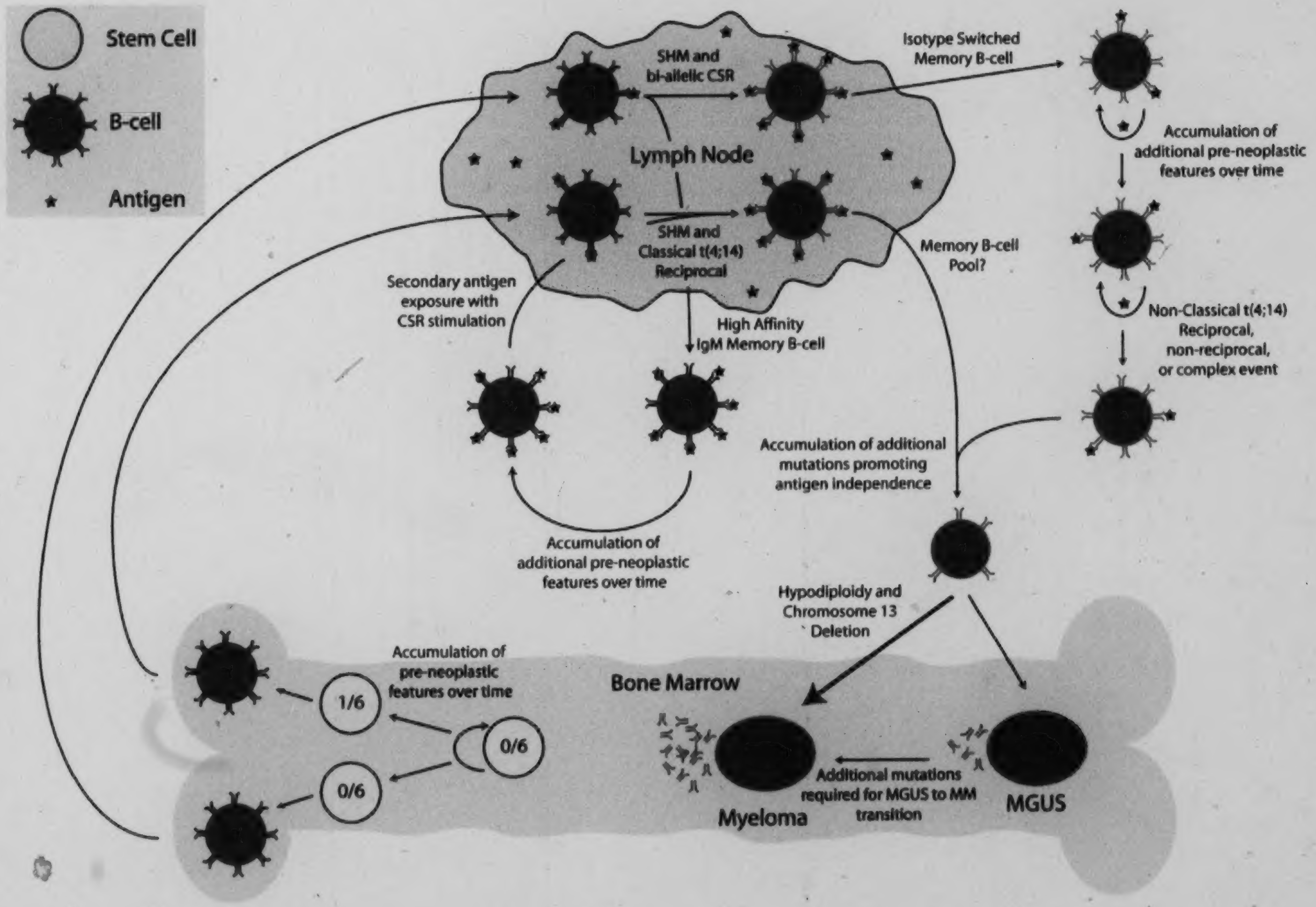

Stem Cell
B-cell
Antigen
SHM and bi-allelic CSR
Lymph Node
SHM and Classical t(4;14) Reciprocal
Isotype Switched Memory B-cell
Accumulation of additional pre-neoplastic features over time
Memory B-cell Pool?
Non-Classical t(4;14) Reciprocal, non-reciprocal, or complex event
Secondary antigen exposure with CSR stimulation
High Affinity IgM Memory B-cell
Accumulation of additional mutations promoting antigen independence
Accumulation of additional pre-neoplastic features over time
Hypodiploidy and Chromosome 13 Deletion
Accumulation of pre-neoplastic features over time
Bone Marrow
1/6
0/6
0/6
Myeloma
MGUS
Additional mutations required for MGUS to MM transition

Figure D.1 – Models of t(4;14) Myelomagenesis

Two different models of t(4;14) mediated myelomagenesis are presented to fit the occurrence of classical and non-classical translocation events. In patients with classical "primary switch" translocations several "pre-neoplastic" events are predicted to exist in the cell in which the translocation occurs. These "pre-neoplastic" features may originate from a subset of haematopoietic stem cells, a mature B-cell that gain the features during development, or an IgM memory B-cell. This "pre-neoplastic" B-cell subsequently encounters antigen (primary or secondary exposure) and undergoes affinity maturation (primary) followed by an illegitimate class switch recombination event (primary or secondary) resulting in t(4;14). Due to the "pre-neoplastic" features in these B-cells the overexpression of MMSET protein variants is tolerated. This B-cell clone may or may not enter the memory B-cell pool, however, at some point this B-cell clone with t(4;14) acquires antigen independence. In patients with non-classical "post-switch strand invasion" translocations a normal B-cell clone encounters antigen and undergoes affinity maturation followed by a legitimate bi-allelic class switch recombination event. This B-cell clone enters the memory B-cell pool and accumulates several "pre-neoplastic" features which allow a subsequent non-classical t(4;14) event and associated overexpression of MMSET protein variants to be tolerated. Subsequently, this B-cell clone with t(4;14) gains antigen independence. At this point both models merge. The transition to the pre-malignant MGUS phase or overt myeloma from either of these B-cell clones is predicted to involve the evolution of a hypodiploid karyotype and an associated chromosome 13 deletion. Finally, in the minority of patients additional mutations such as

TP53 and *PTEN* deletions or *KRAS* mutations are accumulated during the transition from MGUS to overt myeloma.

D.3.1 – Significance of This Work

The body of work documented in this thesis includes several fundamental observations. Although the impact of a body of work is difficult to gauge in such a short time frame, I would suggest the following are the most substantial observations (listed in order of significance):

1) The observation that der(14) is absent from some t(4;14) positive patients, suggesting the translocation is not reciprocal[1].
    A. This suggestion was confirmed in t(4;14) by three independent groups[26,82,112]
    B. More importantly, Fonseca et al. subsequently showed this phenomenon is not limited to t(4;14) but occurs in other recurrent IgH translocations at similar rates[26]. Therefore, this observation significantly impacted the myeloma genetics field at large and forced a change in the previous dogma.
2) The observation that 50% of IgH translocations are mediated by a process that is likely to be independent of the class switch recombination process.
    A. Although this observation is not yet published, it represents a potential paradigm shift in our understanding of the mechanism underlying t(4;14) and the other IgH translocation observed in multiple myeloma.
3) The observation that t(4;14) predicts for a poor prognosis[1].

   A. Although we were the second group to formally publish this observation, in actual fact, it was more or less made simultaneously by three independent groups (Avet-Loiseau, Pilarski, and Fonseca)[1,26,83].

4) The observation that t(4;14) is a poor prognostic factor irrespective of FGFR3 expression suggesting the dysregulation of this gene is not clinically significant[1,122].

5) The observation that transcripts encoding MMSET I, MMSET II, and RE-IIBP are the only universally dysregulated transcripts from 4p16 in t(4;14) myeloma cases, suggesting one or more of these proteins is intimately involved in the myelomagenic program mediated by t(4;14)[122].

6) The observation that t(4;14) mediated myelomagenesis is either a direct result of RE-IIBP overexpression or the indirect result of the overexpression of MMSET I and MMSET II variants since these variants are not functionally equivalent in all t(4;14) positive patients[122].

As outlined above, the work detailed in this thesis has impacted the myeloma research community. We identified one of the worst prognostic markers in multiple myeloma and showed the initially proposed target gene did not impact clinical outcome. Subsequently, we identified several transcripts which are universally dysregulated as a direct result of t(4;14). Moreover, we showed that RE-IIBP is the only protein encoded by these transcripts which is functionally equivalent in all t(4;14) positive patients. Finally, our observation that approximately 50% of t(4;14) translocations originate from a non-classical translocation event (ie. non-CSR mechanism) fundamentally changes our understanding of IgH translocations in myeloma. If the "post-switch strand invasion" model is assumed, the translocation is predicted to occur in a post-switch memory B-cell through a non-CSR mechanism. However, if the "CSR refinement" model is assumed, the translocation must occur before class switch recombination, but importantly, it may even occur before somatic hypermutation or even VDJ recombination. Therefore, in the future it will be essential to determine which non-classical translocation model is correct.

Bibliography

1. Keats JJ, Reiman T, Maxwell CA, Taylor BJ, Larratt LM, Mant MJ, Belch AR, Pilarski LM. In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. Blood. 2003;101:1520-1529

2. Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. Br J Haematol. 2003;121:749-757

3. Gertz MA, Merlini G, Treon SP. Amyloidosis and Waldenstrom's Macroglobulinemia. Hematology. 2004;2004:257-282

4. Alexiou C, Kau RJ, Dietzfelbinger H, Kremer M, Spiess JC, Schratzenstaller B, Arnold W. Extramedullary plasmacytoma: tumor occurrence and therapeutic concepts. Cancer. 1999;85:2305-2314

5. Kyle RA, Therneau TM, Rajkumar SV, Larson DR, Plevak MF, Melton LJ, 3rd. Long-term follow-up of 241 patients with monoclonal gammopathy of undetermined significance: the original Mayo Clinic series 25 years later. Mayo Clin Proc. 2004;79:859-866

6. Kyle RA, Therneau TM, Rajkumar SV, Offord JR, Larson DR, Plevak MF, Melton LJ, III. A Long-Term Study of Prognosis in Monoclonal Gammopathy of Undetermined Significance. N Engl J Med. 2002;346:564-569

7. Goasguen JE, Zandecki M, Mathiot C, Scheiff J-M, Bizet M, Ly-Sunnaram B, Grosbois B, Monconduit M, Michaux J-L, Facon T. Mature plasma cells as indicator of better prognosis in multiple myeloma. New methodology for the assessment of plasma cell morphology. Leukemia Research. 1999;23:1133-1140

8. Greipp PR, Leong T, Bennett JM, Gaillard JP, Klein B, Stewart JA, Oken MM, Kay NE, Van Ness B, Kyle RA. Plasmablastic Morphology---An Independent Prognostic Factor With Clinical and Laboratory Correlates: Eastern Cooperative Oncology Group (ECOG) Myeloma Trial E9486 Report by the ECOG Myeloma Laboratory Group. Blood. 1998;91:2501-2507

9. Rajkumar SV, Fonseca R, Lacy MQ, Witzig TE, Therneau TM, Kyle RA, Litzow MR, Gertz MA, Greipp PR. Plasmablastic morphology is an independent predictor of poor survival after autologous stem-cell transplantation for multiple myeloma. J Clin Oncol. 1999;17:1551-1557

10. Rawstron AC, Owen RG, Davies FE, Johnson RJ, Jones RA, Richards SJ, Evans PA, Child JA, Smith GM, Jack AS, Morgan GJ. Circulating plasma cells in multiple myeloma: characterization and correlation with disease stage. Br J Haematol. 1997;97:46-55

11. Ocqueteau M, Orfao A, Almeida J, Blade J, Gonzalez M, Garcia-Sanz R, Lopez-Berges C, Moro MJ, Hernandez J, Escribano L, Caballero D, Rozman M, San Miguel JF. Immunophenotypic characterization of plasma cells from monoclonal gammopathy of undetermined significance patients. Implications for the differential diagnosis between MGUS and multiple myeloma. Am J Pathol. 1998;152:1655-1665

12. Almeida J, Orfao A, Ocqueteau M, Mateo G, Corral M, Caballero MD, Blade J, Moro MJ, Hernandez J, San Miguel JF. High-sensitive immunophenotyping and DNA ploidy studies for the investigation of minimal residual disease in multiple myeloma. Br J Haematol. 1999;107:121-131

13. Rawstron AC, Fenton JA, Ashcroft J, English A, Jones RA, Richards SJ, Pratt G, Owen R, Davies FE, Child JA, Jack AS, Morgan G. The interleukin-6 receptor alpha-chain (CD126) is expressed by neoplastic but not normal plasma cells. Blood. 2000;96:3880-3886

14. Robillard N, Pellat-Deceunynck C, Bataille R. Phenotypic characterization of the human myeloma cell growth fraction. Blood. 2005:2004-2012-4700

15. Bakkus MH, Heirman C, Van Riet I, Van Camp B, Thielemans K. Evidence that multiple myeloma Ig heavy chain VDJ genes contain somatic mutations but show no intraclonal variation. Blood. 1992;80:2326-2335

16. Billadeau D, Quam L, Thomas W, Kay N, Greipp P, Kyle R, Oken MM, Van Ness B. Detection and quantitation of malignant cells in the peripheral blood of multiple myeloma patients. Blood. 1992;80:1818-1824

17. Bakkus MH, Van Riet I, De Greef C, Van Camp B, Thielemans K. The clonogenic precursor cell in multiple myeloma. Leuk Lymphoma. 1995;18:221-229

18. Bakkus MH, Van Riet I, Van Camp B, Thielemans K. Evidence that the clonogenic cell in multiple myeloma originates from a pre-switched but somatically mutated B cell. Br J Haematol. 1994;87:68-74

19. Billadeau D, Greipp P, Ahmann G, Witzig T, Van Ness B. Detection of B-cells clonally related to the tumor population in multiple myeloma and MGUS. Curr Top Microbiol Immunol. 1995;194:9-16

20. Szczepek AJ, Bergsagel PL, Axelsson L, Brown CB, Belch AR, Pilarski LM. CD34+ cells in the blood of patients with multiple myeloma express CD19 and IgH mRNA and have patient-specific IgH VDJ gene rearrangements. Blood. 1997;89:1824-1833

21. Szczepek AJ, Seeberger K, Wizniak J, Mant MJ, Belch AR, Pilarski LM. A high frequency of circulating B cells share clonotypic Ig heavy-chain VDJ rearrangements with autologous bone marrow plasma cells in multiple myeloma, as measured by single-

cell and in situ reverse transcriptase-polymerase chain reaction. Blood. 1998;92:2844-2855

22. Rasmussen T, Jensen L, Honore L, Andersen H, Johnsen HE. Circulating clonal cells in multiple myeloma do not express CD34 mRNA, as measured by single-cell and real-time RT-PCR assays. Br J Haematol. 1999;107:818-824

23. Rasmussen T, Jensen L, Johnsen HE. The CD19 compartment in myeloma includes a population of clonal cells persistent after high-dose treatment. Leuk Lymphoma. 2002;43:1075-1077

24. Rasmussen T, Lodahl M, Hancke S, Johnsen HE. In multiple myeloma clonotypic CD38- /CD19+ / CD27+ memory B cells recirculate through bone marrow, peripheral blood and lymph nodes. Leuk Lymphoma. 2004;45:1413-1417

25. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature. 2000;408:307-310

26. Fonseca R, Blood E, Rue M, Harrington D, Oken MM, Kyle RA, Dewald GW, Van Ness B, Van Wier SA, Henderson KJ, Bailey RJ, Greipp PR. Clinical and biologic implications of recurrent genomic aberrations in myeloma. Blood. 2003;101:4569-4575

27. Chang H, Qi C, Yi QL, Reece D, Stewart AK. p53 gene deletion detected by fluorescence in situ hybridization is an adverse prognostic factor for patients with multiple myeloma following autologous stem cell transplantation. Blood. 2005;105:358-360

28. Drach J, Ackermann J, Fritz E, Kromer E, Schuster R, Gisslinger H, DeSantis M, Zojer N, Fiegl M, Roka S, Schuster J, Heinz R, Ludwig H, Huber H. Presence of a p53 gene deletion in patients with multiple myeloma predicts for short survival after conventional-dose chemotherapy. Blood. 1998;92:802-809

29. Avet-Loiseau H, Li JY, Godon C, Morineau N, Daviet A, Harousseau JL, Facon T, Bataille R. P53 deletion is not a frequent event in multiple myeloma. Br J Haematol. 1999;106:717-719

30. Neri A, Baldini L, Trecca D, Cro L, Polli E, Maiolo AT. p53 gene mutations in multiple myeloma are associated with advanced forms of malignancy. Blood. 1993;81:128-135

31. Ortega MM, Melo MB, De Souza CA, Lorand-Metze I, Costa FF, Lima CS. A possible role of the P53 gene deletion as a prognostic factor in multiple myeloma. Ann Hematol. 2003;82:405-409

32. Rasmussen T, Kuehl M, Lodahl M, Johnsen HE, Dahl IM. Possible roles for activating RAS mutations in the MGUS to MM transition and in the intramedullary to extramedullary transition in some plasma cell tumors. Blood. 2005;105:317-323

33. Bezieau S, Devilder MC, Avet-Loiseau H, Mellerin MP, Puthier D, Pennarun E, Rapp MJ, Harousseau JL, Moisan JP, Bataille R. High incidence of N and K-Ras activating mutations in multiple myeloma and primary plasma cell leukemia at diagnosis. Hum Mutat. 2001;18:212-224

34. Liu P, Leong T, Quam L, Billadeau D, Kay NE, Greipp P, Kyle RA, Oken MM, Van Ness B. Activating mutations of N- and K-ras in multiple myeloma show different clinical associations: analysis of the Eastern Cooperative Oncology Group Phase III Trial. Blood. 1996;88:2699-2706

35. Chesi M, Brents LA, Ely SA, Bais C, Robbiani DF, Mesri EA, Kuehl WM, Bergsagel PL. Activated fibroblast growth factor receptor 3 is an oncogene that contributes to tumor progression in multiple myeloma. Blood. 2001;97:729-736

36. Trudel S, Li ZH, Wei E, Wiesmann M, Chang H, Chen C, Reece D, Heise C, Stewart AK. CHIR-258, a novel, multi-targeted tyrosine kinase inhibitor for the potential treatment of t(4;14) multiple myeloma. Blood. 2004

37. Goberdhan DCI, Wilson C. PTEN: tumour suppressor, multifunctional growth regulator and more. Hum. Mol. Genet. 2003;12:239R-248

38. Hyun T, Yam A, Pece S, Xie X, Zhang J, Miki T, Gutkind JS, Li W. Loss of PTEN expression leading to high Akt activation in human multiple myelomas. Blood. 2000;96:3560-3568

39. Sakai A, Thieblemont C, Wellmann A, Jaffe ES, Raffeld M. PTEN gene alterations in lymphoid neoplasms. Blood. 1998;92:3410-3415

40. Shi Y, Gera J, Hu L, Hsu JH, Bookstein R, Li W, Lichtenstein A. Enhanced sensitivity of multiple myeloma cells containing PTEN mutations to CCI-779. Cancer Res. 2002;62:5027-5034

41. Ge NL, Rudikoff S. Expression of PTEN in PTEN-deficient multiple myeloma cells abolishes tumor growth in vivo. Oncogene. 2000;19:4091-4095

42. Lewis JP, MacKenzie MR. Non-random chromosomal aberrations associated with multiple myeloma. Hematol Oncol. 1984;2:307-317

43. Debes-Marun CS, Dewald GW, Bryant S, Picken E, Santana-Davila R, Gonzalez-Paz N, Winkler JM, Kyle RA, Gertz MA, Witzig TE, Dispenzieri A, Lacy MQ, Rajkumar SV, Lust JA, Greipp PR, Fonseca R. Chromosome abnormalities clustering and its implications for pathogenesis and prognosis in myeloma. Leukemia. 2003;17:427-436

44. Smadja NV, Bastard C, Brigaudeau C, Leroux D, Fruchart C. Hypodiploidy is a major prognostic factor in multiple myeloma. Blood. 2001;98:2229-2238

45. Lai JL, Zandecki M, Mary JY, Bernardi F, Izydorczyk V, Flactif M, Morel P, Jouet JP, Bauters F, Facon T. Improved cytogenetics in multiple myeloma: a study of 151 patients including 117 patients at diagnosis. Blood. 1995;85:2490-2497

46. Sawyer JR, Waldron JA, Jagannath S, Barlogie B. Cytogenetic findings in 200 patients with multiple myeloma. Cancer Genet Cytogenet. 1995;82:41-49

47. Desikan R, Barlogie B, Sawyer J, Ayers D, Tricot G, Badros A, Zangari M, Munshi NC, Anaissie E, Spoon D, Siegel D, Jagannath S, Vesole D, Epstein J, Shaughnessy J, Fassas A, Lim S, Roberson P, Crowley J. Results of high-dose therapy for 1000 patients with multiple myeloma: durable complete remissions and superior survival in the absence of chromosome 13 abnormalities. Blood. 2000;95:4008-4010

48. Facon T, Avet-Loiseau H, Guillerm G, Moreau P, Genevieve F, Zandecki M, Lai JL, Leleu X, Jouet JP, Bauters F, Harousseau JL, Bataille R, Mary JY. Chromosome 13 abnormalities identified by FISH analysis and serum beta2-microglobulin produce a powerful myeloma staging system for patients receiving high-dose therapy. Blood. 2001;97:1566-1571

49. Fassas AB, Tricot G. Chromosome 13 deletion/hypodiploidy and prognosis in multiple myeloma patients. Leuk Lymphoma. 2004;45:1083-1091

50. Fonseca R, Harrington D, Oken MM, Dewald GW, Bailey RJ, Van Wier SA, Henderson KJ, Blood EA, Rajkumar SV, Kay NE, Van Ness B, Greipp PR. Biological and prognostic significance of interphase fluorescence in situ hybridization detection of chromosome 13 abnormalities (delta13) in multiple myeloma: an eastern cooperative oncology group study. Cancer Res. 2002;62:715-720

51. Kaufmann H, Kromer E, Nosslinger T, Weltermann A, Ackermann J, Reisner R, Bernhart M, Drach J. Both chromosome 13 abnormalities by metaphase cytogenetics and deletion of 13q by interphase FISH only are prognostically relevant in multiple myeloma. Eur J Haematol. 2003;71:179-183

52. Kroger N, Schilling G, Einsele H, Liebisch P, Shimoni A, Nagler A, Perez-Simon JA, San Miguel JF, Kiehl M, Fauser A, Schwerdtfeger R, Wandt H, Sayer HG, Myint H, Klingemann H, Zabelina T, Dierlamm J, Hinke A, Zander AR. Deletion of chromosome band 13q14 as detected by fluorescence in situ hybridization is a prognostic factor in patients with multiple myeloma who are receiving allogeneic dose-reduced stem cell transplantation. Blood. 2004;103:4056-4061

53. Shaughnessy J, Jacobson J, Sawyer J, McCoy J, Fassas A, Zhan F, Bumm K, Epstein J, Anaissie E, Jagannath S, Vesole D, Siegel D, Desikan R, Munshi N, Badros A, Tian E,

Zangari M, Tricot G, Crowley J, Barlogie B. Continuous absence of metaphase-defined cytogenetic abnormalities, especially of chromosome 13 and hypodiploidy, ensures long-term survival in multiple myeloma treated with Total Therapy I: interpretation in the context of global gene expression. Blood. 2003;101:3849-3856

54. Zojer N, Konigsberg R, Ackermann J, Fritz E, Dallinger S, Kromer E, Kaufmann H, Riedl L, Gisslinger H, Schreiber S, Heinz R, Ludwig H, Huber H, Drach J. Deletion of 13q14 remains an independent adverse prognostic variable in multiple myeloma despite its frequent detection by interphase fluorescence in situ hybridization. Blood. 2000;95:1925-1930

55. Tricot G, Barlogie B, Jagannath S, Bracy D, Mattox S, Vesole DH, Naucke S, Sawyer JR. Poor prognosis in multiple myeloma is associated only with partial or complete deletions of chromosome 13 or abnormalities involving 11q and not with other karyotype abnormalities. Blood. 1995;86:4250-4256

56. Bernasconi P, Cavigliano PM, Boni M, Astori C, Calatroni S, Giardini I, Rocca B, Caresana M, Crosetto N, Lazzarino M, Bernasconi C. Long-term follow up with conventional cytogenetics and band 13q14 interphase/metaphase in situ hybridization monitoring in monoclonal gammopathies of undetermined significance. Br J Haematol. 2002;118:545-549

57. Avet-Loiseau H, Li JY, Morineau N, Facon T, Brigaudeau C, Harousseau JL, Grosbois B, Bataille R. Monosomy 13 is associated with the transition of monoclonal gammopathy of undetermined significance to multiple myeloma. Intergroupe Francophone du Myelome. Blood. 1999;94:2583-2589

58. Gutierrez NC, Garcia JL, Hernandez JM, Lumbreras E, Castellanos M, Rasillo A, Mateo G, Perez S, Orfao A, San Miguel Izquierdo JF. Prognostic and biological significance of chromosomal imbalances assessed by comparative genomic hybridization in multiple myeloma. Blood. 2004

59. Liebisch P, Viardot A, Bassermann N, Wendl C, Roth K, Goldschmidt H, Einsele H, Straka C, Stilgenbauer S, Dohner H, Bentz M. Value of comparative genomic hybridization and fluorescence in situ hybridization for molecular diagnostics in multiple myeloma. Br J Haematol. 2003;122:193-201

60. Cigudosa JC, Rao PH, Calasanz MJ, Odero MD, Michaeli J, Jhanwar SC, Chaganti RS. Characterization of nonrandom chromosomal gains and losses in multiple myeloma by comparative genomic hybridization. Blood. 1998;91:3007-3010

61. Almeida J, Orfao A, Mateo G, Ocqueteau M, Garcia-Sanz R, Moro MJ, Hernandez J, Ortega F, Borrego D, Barez A, Mejido M, San Miguel JF. Immunophenotypic and DNA content characteristics of plasma cells in multiple myeloma and monoclonal gammopathy of undetermined significance. Pathol Biol (Paris). 1999;47:119-127

62. San Miguel JF, Garcia-Sanz R, Gonzalez M, Orfao A. DNA cell content studies in multiple myeloma. Leuk Lymphoma. 1996;23:33-41

63. Smadja NV, Leroux D, Soulier J, Dumont S, Arnould C, Taviaux S, Taillemite JL, Bastard C. Further cytogenetic characterization of multiple myeloma confirms that 14q32 translocations are a very rare event in hyperdiploid cases. Genes Chromosomes Cancer. 2003;38:234-239

64. Boersma-Vreugdenhil GR, Peeters T, Bast BJ, Lokhorst HM. Translocation of the IgH locus is nearly ubiquitous in multiple myeloma as detected by immuno-FISH. Blood. 2003;101:1653

65. Avet-Loiseau H, Facon T, Grosbois B, Magrangeas F, Rapp MJ, Harousseau JL, Minvielle S, Bataille R. Oncogenesis of multiple myeloma: 14q32 and 13q chromosomal abnormalities are not randomly distributed, but correlate with natural history, immunological features, and clinical presentation. Blood. 2002;99:2185-2191

66. Nishida K, Tamura A, Nakazawa N, Ueda Y, Abe T, Matsuda F, Kashima K, Taniwaki M. The Ig heavy chain gene is frequently involved in chromosomal translocations in multiple myeloma and plasma cell leukemia as detected by in situ hybridization. Blood. 1997;90:526-534

67. Fonseca R, Barlogie B, Bataille R, Bastard C, Bergsagel PL, Chesi M, Davies FE, Drach J, Greipp PR, Kirsch IR, Kuehl WM, Hernandez JM, Minvielle S, Pilarski LM, Shaughnessy JD, Jr., Stewart AK, Avet-Loiseau H. Genetics and cytogenetics of multiple myeloma: a workshop report. Cancer Res. 2004;64:1546-1558

68. Sawyer JR, Lukacs JL, Thomas EL, Swanson CM, Goosen LS, Sammartino G, Gilliland JC, Munshi NC, Tricot G, Shaughnessy JD, Jr., Barlogie B. Multicolour spectral karyotyping identifies new translocations and a recurring pathway for chromosome loss in multiple myeloma. Br J Haematol. 2001;112:167-174

69. Philip P. Marker chromosome 14q + in multiple myeloma. Hereditas. 1975;80:155-156

70. Liang W, Rowley JD. 14q+ marker chromosomes in multiple myeloma and plasma-cell leukaemia. Lancet. 1978;1:96

71. Wurster-Hill DH, McIntyre OR, Cornwell GG, 3rd. Chromosome studies in myelomatosis. Virchows Arch B Cell Pathol. 1978;29:93-97

72. Venti G, Mecucci C, Donti E, Tabilio A. Translocation t(11;14) and trisomy 11q13----qter in multiple myeloma. Ann Genet. 1984;27:53-55

73. Bergsagel PL, Chesi M, Nardini E, Brents LA, Kirby SL, Kuehl WM. Promiscuous translocations into immunoglobulin heavy chain switch regions in multiple myeloma. Proc Natl Acad Sci U S A. 1996;93:13931-13936

74. Yoshida S, Nakazawa N, Iida S, Hayami Y, Sato S, Wakita A, Shimizu S, Taniwaki M, Ueda R. Detection of MUM1/IRF4-IgH fusion in multiple myeloma. Leukemia. 1999;13:1812-1816

75. Shaughnessy J, Jr., Gabrea A, Qi Y, Brents L, Zhan F, Tian E, Sawyer J, Barlogie B, Bergsagel PL, Kuehl M. Cyclin D3 at 6p21 is dysregulated by recurrent chromosomal translocations to immunoglobulin loci in multiple myeloma. Blood. 2001;98:217-223

76. Hanamura I, Iida S, Akano Y, Hayami Y, Kato M, Miura K, Harada S, Banno S, Wakita A, Kiyoi H, Naoe T, Shimizu S, Sonta SI, Nitta M, Taniwaki M, Ueda R. Ectopic expression of MAFB gene in human myeloma cells carrying (14;20)(q32;q11) chromosomal translocations. Jpn J Cancer Res. 2001;92:638-644

77. Boersma-Vreugdenhil GR, Kuipers J, Van Stralen E, Peeters T, Michaux L, Hagemeijer A, Pearson PL, Clevers HC, Bast BJ. The recurrent translocation t(14;20)(q32;q12) in multiple myeloma results in aberrant expression of MAFB: a molecular and genetic analysis of the chromosomal breakpoint. Br J Haematol. 2004;126:355-363

78. Fonseca R, Debes-Marun CS, Picken EB, Dewald GW, Bryant SC, Winkler JM, Blood E, Oken MM, Santana-Davila R, Gonzalez-Paz N, Kyle RA, Gertz MA, Dispenzieri A, Lacy MQ, Greipp PR. The recurrent IgH translocations are highly associated with nonhyperdiploid variant multiple myeloma. Blood. 2003;102:2562-2567

79. Avet-Loiseau H, Gerson F, Magrangeas F, Minvielle S, Harousseau JL, Bataille R. Rearrangements of the c-myc oncogene are present in 15% of primary human multiple myeloma tumors. Blood. 2001;98:3082-3086

80. Chang H, Sloan S, Li D, Zhuang L, Yi QL, Chen CI, Reece D, Chun K, Keith Stewart A. The t(4;14) is associated with poor prognosis in myeloma patients undergoing autologous stem cell transplant. Br J Haematol. 2004;125:64-68

81. Fonseca R, Oken MM, Greipp PR. The t(4;14)(p16.3;q32) is strongly associated with chromosome 13 abnormalities in both multiple myeloma and monoclonal gammopathy of undetermined significance. Blood. 2001;98:1271-1272

82. Santra M, Zhan F, Tian E, Barlogie B, Shaughnessy J, Jr. A subset of multiple myeloma harboring the t(4;14)(p16;q32) translocation lacks FGFR3 expression but maintains an IGH/MMSET fusion transcript. Blood. 2003;101:2374-2376

83. Moreau P, Facon T, Leleu X, Morineau N, Huyghe P, Harousseau JL, Bataille R, Avet-Loiseau H. Recurrent 14q32 translocations determine the prognosis of multiple

myeloma, especially in patients receiving intensive chemotherapy. Blood. 2002;100:1579-1583

84. Avet-Loiseau H, Facon T, Daviet A, Godon C, Rapp MJ, Harousseau JL, Grosbois B, Bataille R. 14q32 translocations and monosomy 13 observed in monoclonal gammopathy of undetermined significance delineate a multistep process for the oncogenesis of multiple myeloma. Intergroupe Francophone du Myelome. Cancer Res. 1999;59:4546-4550

85. Garand R, Avet-Loiseau H, Accard F, Moreau P, Harousseau JL, Bataille R. t(11;14) and t(4;14) translocations correlated with mature lymphoplasmacytoid and immature morphology, respectively, in multiple myeloma. Leukemia. 2003;17:2032-2035

86. Robillard N, Avet-Loiseau H, Garand R, Moreau P, Pineau D, Rapp MJ, Harousseau JL, Bataille R. CD20 is associated with a small mature plasma cell morphology and t(11;14) in multiple myeloma. Blood. 2003;102:1070-1071

87. Fonseca R, Blood EA, Oken MM, Kyle RA, Dewald GW, Bailey RJ, Van Wier SA, Henderson KJ, Hoyer JD, Harrington D, Kay NE, Van Ness B, Greipp PR. Myeloma and the t(11;14)(q13;q32); evidence for a biologically defined unique subset of patients. Blood. 2002;99:3735-3741

88. Chesi M, Nardini E, Brents LA, Schrock E, Ried T, Kuehl WM, Bergsagel PL. Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. Nat Genet. 1997;16:260-264

89. Fenton JA, Pratt G, Rawstron AC, Sibley K, Rothwell D, Yates Z, Dring A, Richards SJ, Ashcroft AJ, Davies FE, Owen RG, Child JA, Morgan GJ. Genomic characterization of the chromosomal breakpoints of t(4;14) of multiple myeloma suggests more than one possible aetiological mechanism. Oncogene. 2003;22:1103-1113

90. Stec I, Wright TJ, van Ommen GJ, de Boer PA, van Haeringen A, Moorman AF, Altherr MR, den Dunnen JT. WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Hum Mol Genet. 1998;7:1071-1082

91. Chesi M, Nardini E, Lim RS, Smith KD, Kuehl WM, Bergsagel PL. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood. 1998;92:3025-3034

92. Still IH, Vince P, Cowell JK. The third member of the transforming acidic coiled coil-containing gene family, TACC3, maps in 4p16, close to translocation breakpoints in multiple myeloma, and is upregulated in various cancer cell lines. Genomics. 1999;58:165-170

93. Chesi M, Bergsagel PL, Shonukan OO, Martelli ML, Brents LA, Chen T, Schrock E, Ried T, Kuehl WM. Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. Blood. 1998;91:4457-4463

94. Taniwaki M, Nishida K, Ueda Y, Takashima T. Non-random chromosomal rearrangements and their implications in clinical features and outcome of multiple myeloma and plasma cell leukemia. Leuk Lymphoma. 1996;21:25-30

95. Passos-Bueno MR, Wilcox WR, Jabs EW, Sertie AL, Alonso LG, Kitoh H. Clinical spectrum of fibroblast growth factor receptor mutations. Hum Mutat. 1999;14:115-125

96. Richelda R, Ronchetti D, Baldini L, Cro L, Viggiano L, Marzella R, Rocchi M, Otsuki T, Lombardi L, Maiolo AT, Neri A. A novel chromosomal translocation t(4; 14)(p16.3; q32) in multiple myeloma involves the fibroblast growth-factor receptor 3 gene. Blood. 1997;90:4062-4070

97. Sawyer JR, Lukacs JL, Munshi N, Desikan KR, Singhal S, Mehta J, Siegel D, Shaughnessy J, Barlogie B. Identification of new nonrandom translocations in multiple myeloma with multicolor spectral karyotyping. Blood. 1998;92:4269-4278

98. Rao PH, Cigudosa JC, Ning Y, Calasanz MJ, Iida S, Tagawa S, Michaeli J, Klein B, Dalla-Favera R, Jhanwar SC, Ried T, Chaganti RS. Multicolor spectral karyotyping identifies new recurring breakpoints and translocations in multiple myeloma. Blood. 1998;92:1743-1748

99. Nakazawa N, Nishida K, Tamura A, Kobayashi M, Iwai T, Horiike S, Nishigaki H, Otsuki T, Tomiyama Y, Fujii H, Kashima K, Taniwaki M. Interphase detection of t(4;14)(p16.3;q32.3) by in situ hybridization and FGFR3 overexpression in plasma cell malignancies. Cancer Genet Cytogenet. 2000;117:89-96

100. Stewart JP, Thompson A, Santra M, Barlogie B, Lappin TR, Shaughnessy J, Jr. Correlation of TACC3, FGFR3, MMSET and p21 expression with the t(4;14)(p16.3;q32) in multiple myeloma. Br J Haematol. 2004;126:72-76

101. Malgeri U, Baldini L, Perfetti V, Fabris S, Vignarelli MC, Colombo G, Lotti V, Compasso S, Bogni S, Lombardi L, Maiolo AT, Neri A. Detection of t(4;14)(p16.3;q32) chromosomal translocation in multiple myeloma by reverse transcription-polymerase chain reaction analysis of IGH-MMSET fusion transcripts. Cancer Res. 2000;60:4058-4061

102. Avet-Loiseau H, Li JY, Facon T, Brigaudeau C, Morineau N, Maloisel F, Rapp MJ, Talmant P, Trimoreau F, Jaccard A, Harousseau JL, Bataille R. High incidence of translocations t(11;14)(q13;q32) and t(4;14)(p16;q32) in patients with plasma cell malignancies. Cancer Res. 1998;58:5640-5645

103. Finelli P, Fabris S, Zagano S, Baldini L, Intini D, Nobili L, Lombardi L, Maiolo AT, Neri A. Detection of t(4;14)(p16.3;q32) chromosomal translocation in multiple myeloma by double-color fluorescent in situ hybridization. Blood. 1999;94:724-732

104. Avet-Loiseau H, Daviet A, Brigaudeau C, Callet-Bauchu E, Terre C, Lafage-Pochitaloff M, Desangles F, Ramond S, Talmant P, Bataille R. Cytogenetic, interphase, and multicolor fluorescence in situ hybridization analyses in primary plasma cell leukemia: a study of 40 patients at diagnosis, on behalf of the Intergroupe Francophone du Myelome and the Groupe Francais de Cytogenetique Hematologique. Blood. 2001;97:822-825

105. Matsuda F, Shin EK, Nagaoka H, Matsumura R, Haino M, Fukita Y, Taka-ishi S, Imai T, Riley JH, Anand R, et al. Structure and physical map of 64 variable segments in the 3'0.8-megabase region of the human immunoglobulin heavy-chain locus. Nat Genet. 1993;3:88-94

106. Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, Honjo T. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J Exp Med. 1998;188:2151-2162

107. Cook GP, Tomlinson IM, Walter G, Riethman H, Carter NP, Buluwela L, Winter G, Rabbitts TH. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. Nat Genet. 1994;7:162-168

108. Fonseca R, Bailey RJ, Ahmann GJ, Rajkumar SV, Hoyer JD, Lust JA, Kyle RA, Gertz MA, Greipp PR, Dewald GW. Genomic abnormalities in monoclonal gammopathy of undetermined significance. Blood. 2002;100:1417-1424

109. Gabrea A, Bergsagel PL, Chesi M, Shou Y, Kuehl WM. Insertion of excised IgH switch sequences causes overexpression of cyclin D1 in a myeloma tumor cell. Mol Cell. 1999;3:119-123

110. Chang H, Sloan S, Li D, Patterson B. Genomic aberrations in plasma cell leukemia shown by interphase fluorescence in situ hybridization. Cancer Genet Cytogenet. 2005;156:150-153

111. Chang H, Sloan S, Li D, Keith Stewart A. Multiple myeloma involving central nervous system: high frequency of chromosome 17p13.1 (p53) deletions. Br J Haematol. 2004;127:280-284

112. Fabris S, Agnelli L, Mattioli M, Baldini L, Ronchetti D, Morabito F, Verdelli D, Nobili L, Intini D, Callea V, Stelitano C, Lombardi L, Neri A. Characterization of oncogene dysregulation in multiple myeloma by combined FISH and DNA microarray analyses. Genes Chromosomes Cancer. 2005;42:117-127

113. Fabris S, Storlazzi CT, Baldini L, Nobili L, Lombardi L, Maiolo AT, Rocchi M, Neri A. Heterogeneous pattern of chromosomal breakpoints involving the MYC locus in multiple myeloma. Genes Chromosomes Cancer. 2003;37:261-269

114. Wuilleme S, Robillard N, Lode L, Magrangeas F, Beris H, Harousseau JL, Proffitt J, Minvielle S, Avet-Loiseau H. Ploidy, as detected by fluorescence in situ hybridization, defines different subgroups in multiple myeloma. Leukemia. 2005;19:275-278

115. Zhan F, Hardin J, Kordsmeier B, Bumm K, Zheng M, Tian E, Sanderson R, Yang Y, Wilson C, Zangari M, Anaissie E, Morris C, Muwalla F, van Rhee F, Fassas A, Crowley J, Tricot G, Barlogie B, Shaughnessy J, Jr. Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. Blood. 2002;99:1745-1757

116. Munshi NC, Hideshima T, Carrasco D, Shammas M, Auclair D, Davies F, Mitsiades N, Mitsiades C, Kim RS, Li C, Rajkumar SV, Fonseca R, Bergsagel L, Chauhan D, Anderson KC. Identification of genes modulated in multiple myeloma using genetically identical twin samples. Blood. 2004;103:1799-1806

117. Dring AM, Davies FE, Fenton JA, Roddam PL, Scott K, Gonzalez D, Rollinson S, Rawstron AC, Rees-Unwin KS, Li C, Munshi NC, Anderson KC, Morgan GJ. A global expression-based analysis of the consequences of the t(4;14) translocation in myeloma. Clin Cancer Res. 2004;10:5692-5701

118. Soverini S, Terragna C, Testoni N, Ruggeri D, Tosi P, Zamagni E, Cellini C, Cavo M, Baccarani M, Tura S, Martinelli G. Novel mutation and RNA splice variant of fibroblast growth factor receptor 3 in multiple myeloma patients at diagnosis. Haematologica. 2002;87:1036-1040

119. Sibley K, Fenton JA, Dring AM, Ashcroft AJ, Rawstron AC, Morgan GJ. A molecular study of the t(4;14) in multiple myeloma. Br J Haematol. 2002;118:514-520

120. Rasmussen T, Hudlebusch HR, Knudsen LM, Johnsen HE. FGFR3 dysregulation in multiple myeloma: frequency and prognostic relevance. Br J Haematol. 2002;117:626-628

121. Rasmussen T, Theilgaard-Monch K, Hudlebusch HR, Lodahl M, Johnsen HE, Dahl IM. Occurrence of dysregulated oncogenes in primary plasma cells representing consecutive stages of myeloma pathogenesis: indications for different disease entities. Br J Haematol. 2003;123:253-262

122. Keats JJ, Maxwell CA, Taylor BJ, Hendzel MJ, Chesi M, Bergsagel PL, Larratt LM, Mant MJ, Reiman T, Belch AR, Pilarski LM. Overexpression of transcripts originating from the MMSET Locus characterizes all t(4;14)(p16;q32) positive multiple myeloma patients. Blood. 2005;105:4060-4069

123. Kyle RA, Lust JA. Monoclonal gammopathies of undetermined significance. Semin Hematol. 1989;26:176-200

124. Kyle RA, Greipp PR. Smoldering multiple myeloma. N Engl J Med. 1980;302:1347-1349

125. Oken MM, Leong T, Lenhard RE, Jr., Greipp PR, Kay NE, Van Ness B, Keimowitz RM, Kyle RA. The addition of interferon or high dose cyclophosphamide to standard chemotherapy in the treatment of patients with multiple myeloma: phase III Eastern Cooperative Oncology Group Clinical Trial EST 9486. Cancer. 1999;86:957-968

126. Still IH, Vettaikkorumakankauv AK, DiMatteo A, Liang P. Structure-function evolution of the Transforming acidic coiled coil genes revealed by analysis of phylogenetically diverse organisms. BMC Evol Biol. 2004;4:16

127. Pebusque MJ, Coulier F, Birnbaum D, Pontarotti P. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol. 1998;15:1145-1159

128. Vienne A, Rasmussen J, Abi-Rached L, Pontarotti P, Gilles A. Systematic phylogenomic evidence of en bloc duplication of the ancestral 8p11.21-8p21.3-like region. Mol Biol Evol. 2003;20:1290-1298

129. Li Z, Zhu YX, Plowright EE, Bergsagel PL, Chesi M, Patterson B, Hawley TS, Hawley RG, Stewart AK. The myeloma-associated oncogene fibroblast growth factor receptor 3 is transforming in hematopoietic cells. Blood. 2001;97:2413-2419

130. Plowright EE, Li Z, Bergsagel PL, Chesi M, Barber DL, Branch DR, Hawley RG, Stewart AK. Ectopic expression of fibroblast growth factor receptor 3 promotes myeloma cell proliferation and prevents apoptosis. Blood. 2000;95:992-998

131. Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics. 2001;17:1244-1245

132. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-4680

133. Still IH, Hamilton M, Vince P, Wolfman A, Cowell JK. Cloning of TACC1, an embryonically expressed, potentially transforming coiled coil containing gene, from the 8p11 breast cancer amplicon. Oncogene. 1999;18:4032-4038

134. Chen HM, Schmeichel KL, Mian IS, Lelievre S, Petersen OW, Bissell MJ. AZU-1: a candidate breast tumor suppressor and biomarker for tumor progression. Mol Biol Cell. 2000;11:1357-1367

135. Sadek CM, Pelto-Huikko M, Tujague M, Steffensen KR, Wennerholm M, Gustafsson JA. TACC3 expression is tightly regulated during early differentiation. Gene Expr Patterns. 2003;3:203-211

136. Piekorz RP, Hoffmeyer A, Duntsch CD, McKay C, Nakajima H, Sexl V, Snyder L, Rehg J, Ihle JN. The centrosomal protein TACC3 is essential for hematopoietic stem cell function and genetically interfaces with p53-regulated apoptosis. Embo J. 2002;21:653-664

137. Gergely F, Karlsson C, Still I, Cowell J, Kilmartin J, Raff JW. The TACC domain identifies a family of centrosomal proteins that can interact with microtubules. Proc Natl Acad Sci U S A. 2000;97:14352-14357

138. McKeveney PJ, Hodges VM, Mullan RN, Maxwell P, Simpson D, Thompson A, Winter PC, Lappin TR, Maxwell AP. Characterization and localization of expression of an erythropoietin-induced gene, ERIC-1/TACC3, identified in erythroid precursor cells. Br J Haematol. 2001;112:1016-1024

139. Gergely F, Draviam VM, Raff JW. The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells. Genes Dev. 2003;17:336-341

140. Schuendeln MM, Piekorz RP, Wichmann C, Lee Y, McKinnon PJ, Boyd K, Takahashi Y, Ihle JN. The centrosomal, putative tumor suppressor protein TACC2 is dispensable for normal development, and deficiency does not lead to cancer. Mol Cell Biol. 2004;24:6403-6409

141. Gergely F, Kidd D, Jeffers K, Wakefield JG, Raff JW. D-TACC: a novel centrosomal protein required for normal spindle function in the early Drosophila embryo. Embo J. 2000;19:241-252

142. Garriga-Canut M, Orkin SH. Transforming acidic coiled-coil protein 3 (TACC3) controls friend of GATA-1 (FOG-1) subcellular localization and regulates the association between GATA-1 and FOG-1 during hematopoiesis. J Biol Chem. 2004;279:23597-23605

143. Simpson RJ, Yi Lee SH, Bartle N, Sum EY, Visvader JE, Matthews JM, Mackay JP, Crossley M. A classic zinc finger from friend of GATA mediates an interaction with the coiled-coil of transforming acidic coiled-coil 3. J Biol Chem. 2004;279:39789-39797

144. Gangisetty O, Lauffart B, Sondarva GV, Chelsea DM, Still IH. The transforming acidic coiled coil proteins interact with nuclear histone acetyltransferases. Oncogene. 2004;23:2559-2563

145. Dib A, Adelaide J, Chaffanet M, Imbert A, Le Paslier D, Jacquemier J, Gaudray P, Theillet C, Birnbaum D, Pebusque MJ. Characterization of the region of the short arm of chromosome 8 amplified in breast carcinoma. Oncogene. 1995;10:995-1001

146. Keegan K, Johnson DE, Williams LT, Hayman MJ. Isolation of an additional member of the fibroblast growth factor receptor family, FGFR-3. Proc Natl Acad Sci U S A. 1991;88:1095-1099

147. Thompson LM, Plummer S, Schalling M, Altherr MR, Gusella JF, Housman DE, Wasmuth JJ. A gene encoding a fibroblast growth factor receptor isolated from the Huntington disease gene region of human chromosome 4. Genomics. 1991;11:1133-1142

148. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, et al. A polymorphic DNA marker genetically linked to Huntington's disease. Nature. 1983;306:234-238

149. Gusella JF, Tanzi RE, Bader PI, Phelan MC, Stevenson R, Hayden MR, Hofman KJ, Faryniarz AG, Gibbons K. Deletion of Huntington's disease-linked G8 (D4S10) locus in Wolf-Hirschhorn syndrome. Nature. 1985;318:75-78

150. Gilliam TC, Tanzi RE, Haines JL, Bonner TI, Faryniarz AG, Hobbs WJ, MacDonald ME, Cheng SV, Folstein SE, Conneally PM, et al. Localization of the Huntington's disease gene to a small segment of chromosome 4 flanked by D4S10 and the telomere. Cell. 1987;50:565-571

151. Velinov M, Slaugenhaupt SA, Stoilov I, Scott CI, Jr., Gusella JF, Tsipouras P. The gene for achondroplasia maps to the telomeric region of chromosome 4p. Nat Genet. 1994;6:314-317

152. Le Merrer M, Rousseau F, Legeai-Mallet L, Landais JC, Pelet A, Bonaventure J, Sanak M, Weissenbach J, Stoll C, Munnich A, et al. A gene for achondroplasia-hypochondroplasia maps to chromosome 4p. Nat Genet. 1994;6:318-321

153. Francomano CA, Ortiz de Luna RI, Hefferon TW, Bellus GA, Turner CE, Taylor E, Meyers DA, Blanton SH, Murray JC, McIntosh I, et al. Localization of the achondroplasia gene to the distal 2.5 Mb of human chromosome 4p. Hum Mol Genet. 1994;3:787-792

154. Shiang R, Thompson LM, Zhu YZ, Church DM, Fielder TJ, Bocian M, Winokur ST, Wasmuth JJ. Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. Cell. 1994;78:335-342

155. Rousseau F, Bonaventure J, Legeai-Mallet L, Pelet A, Rozet JM, Maroteaux P, Le Merrer M, Munnich A. Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. Nature. 1994;371:252-254

156. Vajo Z, Francomano CA, Wilkin DJ. The molecular and genetic basis of fibroblast growth factor receptor 3 disorders: the achondroplasia family of skeletal dysplasias,

Muenke craniosynostosis, and Crouzon syndrome with acanthosis nigricans. Endocr Rev. 2000;21:23-39

157. Monsonego-Ornan E, Adar R, Feferman T, Segev O, Yayon A. The transmembrane mutation G380R in fibroblast growth factor receptor 3 uncouples ligand-mediated receptor activation from down-regulation. Mol Cell Biol. 2000;20:516-522

158. Adar R, Monsonego-Ornan E, David P, Yayon A. Differential activation of cysteine-substitution mutants of fibroblast growth factor receptor 3 is determined by cysteine localization. J Bone Miner Res. 2002;17:860-868

159. Bellus GA, Spector EB, Speiser PW, Weaver CA, Garber AT, Bryke CR, Israel J, Rosengren SS, Webster MK, Donoghue DJ, Francomano CA. Distinct missense mutations of the FGFR3 lys650 codon modulate receptor kinase activation and the severity of the skeletal dysplasia phenotype. Am J Hum Genet. 2000;67:1411-1421

160. Cappellen D, De Oliveira C, Ricol D, de Medina S, Bourdin J, Sastre-Garau X, Chopin D, Thiery JP, Radvanyi F. Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. Nat Genet. 1999;23:18-20

161. Sibley K, Cuthbert-Heavens D, Knowles MA. Loss of heterozygosity at 4p16.3 and mutation of FGFR3 in transitional cell carcinoma. Oncogene. 2001;20:686-691

162. Kimura T, Suzuki H, Ohashi T, Asano K, Kiyota H, Eto Y. The incidence of thanatophoric dysplasia mutations in FGFR3 gene is higher in low-grade or superficial bladder carcinomas. Cancer. 2001;92:2555-2561

163. Bakkar AA, Wallerand H, Radvanyi F, Lahaye JB, Pissard S, Lecerf L, Kouyoumdjian JC, Abbou CC, Pairon JC, Jaurand MC, Thiery JP, Chopin DK, de Medina SG. FGFR3 and TP53 gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder. Cancer Res. 2003;63:8108-8112

164. Rieger-Christ KM, Mourtzinos A, Lee PJ, Zagha RM, Cain J, Silverman M, Libertino JA, Summerhayes IC. Identification of fibroblast growth factor receptor 3 mutations in urine sediment DNA samples complements cytology in bladder tumor detection. Cancer. 2003;98:737-744

165. van Rhijn BW, Lurkin I, Radvanyi F, Kirkels WJ, van der Kwast TH, Zwarthoff EC. The fibroblast growth factor receptor 3 (FGFR3) mutation is a strong indicator of superficial bladder cancer with low recurrence rate. Cancer Res. 2001;61:1265-1268

166. Wu R, Connolly D, Ngelangel C, Bosch FX, Munoz N, Cho KR. Somatic mutations of fibroblast growth factor receptor 3 (FGFR3) are uncommon in carcinomas of the uterine cervix. Oncogene. 2000;19:5543-5546

167. Sibley K, Stern P, Knowles MA. Frequency of fibroblast growth factor receptor 3 mutations in sporadic tumours. Oncogene. 2001;20:4416-4418

168. Fracchiolla NS, Luminari S, Baldini L, Lombardi L, Maiolo AT, Neri A. FGFR3 gene mutations associated with human skeletal disorders occur rarely in multiple myeloma. Blood. 1998;92:2987-2989

169. Intini D, Baldini L, Fabris S, Lombardi L, Ciceri G, Maiolo AT, Neri A. Analysis of FGFR3 gene mutations in multiple myeloma patients with t(4;14). Br J Haematol. 2001;114:362-364

170. Intini D, Baldini L, Lombardi L, Neri A. A novel mutation involving the carboxy terminal region of the FGFR3 gene in a multiple myeloma patient with t(4;14). Leukemia. 2002;16:1201-1202

171. Onwuazor ON, Wen XY, Wang DY, Zhuang L, Masih-Khan E, Claudio J, Barlogie B, Shaughnessy JD, Jr., Stewart AK. Mutation, SNP, and isoform analysis of fibroblast growth factor receptor 3 (FGFR3) in 150 newly diagnosed multiple myeloma patients. Blood. 2003;102:772-773

172. Winterpacht A, Hilbert K, Stelzer C, Schweikardt T, Decker H, Segerer H, Spranger J, Zabel B. A novel mutation in FGFR-3 disrupts a putative N-glycosylation site and results in hypochondroplasia. Physiol Genomics. 2000;2:9-12

173. Jang JH, Shin KH, Park JG. Mutations in fibroblast growth factor receptor 2 and fibroblast growth factor receptor 3 genes associated with human gastric and colorectal cancers. Cancer Res. 2001;61:3541-3543

174. Ronchetti D, Greco A, Compasso S, Colombo G, Dell'Era P, Otsuki T, Lombardi L, Neri A. Deregulated FGFR3 mutants in multiple myeloma cell lines with t(4;14): comparative analysis of Y373C, K650E and the novel G384D mutations. Oncogene. 2001;20:3553-3562

175. Webster MK, D'Avis PY, Robertson SC, Donoghue DJ. Profound ligand-independent kinase activation of fibroblast growth factor receptor 3 by the activation loop mutation responsible for a lethal skeletal dysplasia, thanatophoric dysplasia type II. Mol Cell Biol. 1996;16:4081-4087

176. Webster MK, Donoghue DJ. Enhanced signaling and morphological transformation by a membrane-localized derivative of the fibroblast growth factor receptor 3 kinase domain. Mol Cell Biol. 1997;17:5739-5747

177. Affer M, Robbiani DF, Chesi M, Bergsagel PL. Ectopic Fibroblast Growth Factor Receptor 3 (FGFR3) Expression Mediated by Isotype Switch Recombination. Blood. 2004;104:25a

178. Schlickum S, Moghekar A, Simpson JC, Steglich C, O'Brien RJ, Winterpacht A, Endele SU. LETM1, a gene deleted in Wolf-Hirschhorn syndrome, encodes an evolutionarily conserved mitochondrial protein. Genomics. 2004;83:254-261

179. Nowikovsky K, Froschauer EM, Zsurka G, Samaj J, Reipert S, Kolisek M, Wiesenberger G, Schweyen RJ. The LETM1/YOL027 gene family encodes a factor of the mitochondrial K+ homeostasis with a potential role in the Wolf-Hirschhorn syndrome. J Biol Chem. 2004;279:30307-30315

180. Brown J, Jawad M, Twigg SR, Saracoglu K, Sauerbrey A, Thomas AE, Eils R, Harbott J, Kearney L. A cryptic t(5;11)(q35;p15.5) in 2 children with acute myeloid leukemia with apparently normal karyotypes, identified by a multiplex fluorescence in situ hybridization telomere assay. Blood. 2002;99:2526-2531

181. Jaju RJ, Fidler C, Haas OA, Strickson AJ, Watkins F, Clark K, Cross NC, Cheng JF, Aplan PD, Kearney L, Boultwood J, Wainscoat JS. A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. Blood. 2001;98:1264-1267

182. La Starza R, Gorello P, Rosati R, Riezzo A, Veronese A, Ferrazzi E, Martelli MF, Negrini M, Mecucci C. Cryptic insertion producing two NUP98/NSD1 chimeric transcripts in adult refractory anemia with an excess of blasts. Genes Chromosomes Cancer. 2004;41:395-399

183. Panarello C, Rosanda C, Morerio C. Cryptic translocation t(5;11)(q35;p15.5) with involvement of the NSD1 and NUP98 genes without 5q deletion in childhood acute myeloid leukemia. Genes Chromosomes Cancer. 2002;35:277-281

184. Rosati R, La Starza R, Veronese A, Aventin A, Schwienbacher C, Vallespi T, Negrini M, Martelli MF, Mecucci C. NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). Blood. 2002;99:3857-3860

185. Tripoulas N, LaJeunesse D, Gildea J, Shearn A. The Drosophila ash1 gene product, which is localized at specific sites on polytene chromosomes, contains a SET domain and a PHD finger. Genetics. 1996;143:913-928

186. Pribill I, Barnes GT, Chen J, Church D, Buckler A, Baxendale S, Bates GP, Lehrach H, Gusella MJ, Duyao MP, Ambrose CM, Gusella JF, MacDonald ME. Exon trapping and sequence-based methods of gene finding in transcript mapping of human 4p16.3. Somat Cell Mol Genet. 1997;23:413-427

187. Wright TJ, Ricke DO, Denison K, Abmayr S, Cotter PD, Hirschhorn K, Keinanen M, McDonald-McGinn D, Somer M, Spinner N, Yang-Feng T, Zackai E, Altherr MR. A transcript map of the newly defined 165 kb Wolf-Hirschhorn syndrome critical region. Hum Mol Genet. 1997;6:317-324

188. Garlisi CG, Uss AS, Xiao H, Tian F, Sheridan KE, Wang L, Motasim Billah M, Egan RW, Stranick KS, Umland SP. A unique mRNA initiated within a middle intron of WHSC1/MMSET encodes a DNA binding protein that suppresses human IL-5 transcription. Am J Respir Cell Mol Biol. 2001;24:90-98

189. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A. 1998;95:5857-5864

190. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. Nucleic Acids Res. 2004;32 Database issue:D142-144

191. Chesi M, Naik K, Robbiani DF, Affer M, Nickerson HD, Bergsagel PL. Dysregulation of MMSET Is Tumorigenic In-Vivo. Blood. 2004;104:24a

192. Stec I, Nagl SB, van Ommen GJ, den Dunnen JT. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS Lett. 2000;473:1-5

193. Kouzarides T. Histone methylation in transcriptional control. Curr Opin Genet Dev. 2002;12:198-209

194. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. 1994;13:3822-3831

195. Stassen MJ, Bailey D, Nelson S, Chinwalla V, Harte PJ. The Drosophila trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. Mechanisms of Development. 1995;52:209-223

196. Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 2004;18:1251-1262

197. Strahl BD, Grant PA, Briggs SD, Sun Z-W, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF, Allis CD. Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. Mol. Cell. Biol. 2002;22:1298-1306

198. Beisel C, Imhof A, Greene J, Kremmer E, Sauer F. Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature. 2002;419:857-862

199. Byrd KN, Shearn A. ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. PNAS. 2003;100:11535-11540

200. Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. Embo J. 2003;22:3153-3163

201. Wu Y, Berends MJ, Mensink RG, Kempinga C, Sijmons RH, van Der Zee AG, Hollema H, Kleibeuker JH, Buys CH, Hofstra RM. Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. Am J Hum Genet. 1999;65:1291-1298

202. Berends MJ, Wu Y, Sijmons RH, Mensink RG, van der Sluis T, Hordijk-Hos JM, de Vries EG, Hollema H, Karrenbeld A, Buys CH, van der Zee AG, Hofstra RM, Kleibeuker JH. Molecular and clinical characteristics of MSH6 variants: an analysis of 25 index carriers of a germline variant. Am J Hum Genet. 2002;70:26-37

203. Shirohzu H, Kubota T, Kumazawa A, Sado T, Chijiwa T, Inagaki K, Suetake I, Tajima S, Wakui K, Miki Y, Hayashi M, Fukushima Y, Sasaki H. Three novel DNMT3B mutations in Japanese patients with ICF syndrome. Am J Med Genet. 2002;112:31-37

204. Kurotaki N, Harada N, Shimokawa O, Miyake N, Kawame H, Uetake K, Makita Y, Kondoh T, Ogata T, Hasegawa T, Nagai T, Ozaki T, Touyama M, Shenhav R, Ohashi H, Medne L, Shiihara T, Ohtsu S, Kato Z, Okamoto N, Nishimoto J, Lev D, Miyoshi Y, Ishikiriyama S, Sonoda T, Sakazume S, Fukushima Y, Kurosawa K, Cheng JF, Yoshiura K, Ohta T, Kishino T, Niikawa N, Matsumoto N. Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion. Hum Mutat. 2003;22:378-387

205. Qiu C, Sawada K, Zhang X, Cheng X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol. 2002;9:217-224

206. Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem. 2004;279:25447-25454

207. Chen T, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol. 2004;24:9048-9058

208. Slater LM, Allen MD, Bycroft M. Structural variation in PWWP domains. J Mol Biol. 2003;330:571-576

209. Grasser KD. Chromatin-associated HMGA and HMGB proteins: versatile co-regulators of DNA-dependent processes. Plant Mol Biol. 2003;53:281-295

210. Aasland R, Gibson TJ, Stewart AF. The PHD finger: Implications for chromatin-mediated transcriptional regulation. Trends in Biochemical Sciences. 1995;20:56-59

211. Eberharter A, Vetter I, Ferreira R, Becker PB. ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. Embo J. 2004;23:4029-4039

212. Ragvin A, Valvatne H, Erdal S, Arskog V, Tufteland KR, Breen K, Oyan AM, Eberharter A, Gibson TJ, Becker PB, Aasland R. Nucleosome Binding by the Bromodomain and PHD Finger of the Transcriptional Cofactor p300. Journal of Molecular Biology. 2004;337:773-788

213. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, Baird CL, Zhu H, Field SJ, Lessnick SL. The PHD Finger of the Chromatin-Associated Protein ING2 Functions as a Nuclear Phosphoinositide Receptor. Cell. 2003;114:99-111

214. Lu Z, Xu S, Joazeiro C, Cobb MH, Hunter T. The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2. Mol Cell. 2002;9:945-956

215. Coscoy L, Ganem D. PHD domains and E3 ubiquitin ligases: viruses make the connection. Trends Cell Biol. 2003;13:7-12

216. Aravind L, Iyer LM, Koonin EV. Scores of RINGS but no PHDs in ubiquitin signaling. Cell Cycle. 2003;2:123-126

217. Scheel H, Hofmann K. No evidence for PHD fingers as ubiquitin ligases. Trends Cell Biol. 2003;13:285-287; author reply 287-288

218. Matthews JM, Sunde M. Zinc fingers--folds for many occasions. IUBMB Life. 2002;54:351-355

219. Jackson PK, Eldridge AG, Freed E, Furstenthal L, Hsu JY, Kaiser BK, Reimann JDR. The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. Trends in Cell Biology. 2000;10:429-439

220. Melchior L, Schwartz M, Duno M. dHPLC Screening of the NSD1 gene Identifies Nine Novel Mutations - Summary of the first 100 Sotos Syndrome Mutations. Ann Hum Genet. 2005;69:222-226

221. Douglas J, Hanks S, Temple IK, Davies S, Murray A, Upadhyaya M, Tomkins S, Hughes HE, Cole TR, Rahman N. NSD1 mutations are the major cause of Sotos

syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. Am J Hum Genet. 2003;72:132-143

222. Rio M, Clech L, Amiel J, Faivre L, Lyonnet S, Le Merrer M, Odent S, Lacombe D, Edery P, Brauner R, Raoul O, Gosset P, Prieur M, Vekemans M, Munnich A, Colleaux L, Cormier-Daire V. Spectrum of NSD1 mutations in Sotos and Weaver syndromes. J Med Genet. 2003;40:436-440

223. Wright TJ, Costa JL, Naranjo C, Francis-West P, Altherr MR. Comparative analysis of a novel gene from the Wolf-Hirschhorn/Pitt-Rogers-Danks syndrome critical region. Genomics. 1999;59:203-212

224. Mariotti M, Manganini M, Maier JA. Modulation of WHSC2 expression in human endothelial cells. FEBS Lett. 2000;487:166-170

225. Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell. 1999;97:41-51

226. Yamaguchi Y, Inukai N, Narita T, Wada T, Handa H. Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol. 2002;22:2918-2927

227. Narita T, Yamaguchi Y, Yano K, Sugimoto S, Chanarat S, Wada T, Kim DK, Hasegawa J, Omori M, Inukai N, Endoh M, Yamada T, Handa H. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. Mol Cell Biol. 2003;23:1863-1873

228. Yamaguchi Y, Filipovska J, Yano K, Furuya A, Inukai N, Narita T, Wada T, Sugimoto S, Konarska MM, Handa H. Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science. 2001;293:124-127

229. Kyle RA, Rajkumar SV. Monoclonal gammopathies of undetermined significance. Hematol Oncol Clin North Am. 1999;13:1181-1202

230. Hussein M. Multiple myeloma: an overview of diagnosis and management. Cleve Clin J Med. 1994;61:285-298

231. Hallek M, Bergsagel PL, Anderson KC. Multiple myeloma: increasing evidence for a multistep transformation process. Blood. 1998;91:3-21

232. Bergsagel PL, Kuehl WM. Chromosome translocations in multiple myeloma. Oncogene. 2001;20:5611-5622

233. Fenton JA, Pratt G, Rothwell DG, Rawstron AC, Morgan GJ. Translocation t(11;14) in multiple myeloma: Analysis of translocation breakpoints on der(11) and der(14) chromosomes suggests complex molecular mechanisms of recombination. Genes Chromosomes Cancer. 2004;39:151-155

234. Perfetti V, Coluccia AM, Intini D, Malgeri U, Vignarelli MC, Casarini S, Merlini G, Neri A. Translocation T(4;14)(p16.3;q32) is a recurrent genetic lesion in primary amyloidosis. Am J Pathol. 2001;158:1599-1603

235. Browman GP, Bergsagel D, Sicheri D, O'Reilly S, Wilson KS, Rubin S, Belch A, Shustik C, Barr R, Walker I, et al. Randomized trial of interferon maintenance in multiple myeloma: a study of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol. 1995;13:2354-2360

236. Intini D, Fabris S, Storlazzi T, Otsuki T, Ciceri G, Verdelli D, Lombardi L, Rocchi M, Neri A. Identification of a novel IGH-MMSET fusion transcript in a human myeloma cell line with the t(4;14)(p16.3;q32) chromosomal translocation. Br J Haematol. 2004;126:437-439

237. Schneider R, Bannister AJ, Kouzarides T. Unsafe SETs: histone lysine methyltransferases and cancer. Trends Biochem Sci. 2002;27:396-402

238. Chen D, Huang S. Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. J Cell Biol. 2001;153:169-176

239. Phair RD, Gorski SA, Misteli T. Measurement of dynamic protein binding to chromatin in vivo, using photobleaching microscopy. Methods Enzymol. 2004;375:393-414

240. Lippincott-Schwartz J, Snapp E, Kenworthy A. Studying protein dynamics in living cells. Nat Rev Mol Cell Biol. 2001;2:444-456

241. Nakamura T, Blechman J, Tada S, Rozovskaia T, Itoyama T, Bullrich F, Mazo A, Croce CM, Geiger B, Canaani E. huASH1 protein, a putative transcription factor encoded by a human homologue of the Drosophila ash1 gene, localizes to both nuclei and cell-cell tight junctions. PNAS. 2000;97:7284-7289

242. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun Z-W, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000;406:593-599

243. Ronchetti D, Bogni S, Finelli P, Lombardi L, Otsuki T, Maiolo AT, Neri A. Characterization of the t(4;14)(p16.3;q32) in the KMS-18 multiple myeloma cell line. Leukemia. 2001;15:864-865

244. Zarrin AA, Alt FW, Chaudhuri J, Stokes N, Kaushal D, Du Pasquier L, Tian M. An evolutionarily conserved target motif for immunoglobulin class-switch recombination. Nat Immunol. 2004;5:1275-1281

245. Zarrin AA, Tian M, Wang J, Borjeson T, Alt FW. Influence of switch region length on immunoglobulin class switch recombination. PNAS. 2005;102:2466-2470

246. Kovalchuk AL, Esa A, Coleman AE, Park SS, Ried T, Cremer CC, Janz S. Translocation remodeling in the primary BALB/c plasmacytoma TEPC 3610. Genes Chromosomes Cancer. 2001;30:283-291

247. Sonoki T, Willis TG, Oscier DG, Karran EL, Siebert R, Dyer MJ. Rapid amplification of immunoglobulin heavy chain switch (IGHS) translocation breakpoints using long-distance inverse PCR. Leukemia. 2004;18:2026-2031

248. Otsuki T, Yata K, Sakaguchi H, Uno M, Fujii T, Wada H, Sugihara T, Ueki A. IL-10 in myeloma cells. Leuk Lymphoma. 2002;43:969-974

249. Gazdar AF, Oie HK, Kirsch IR, Hollis GF. Establishment and characterization of a human plasma cell myeloma culture having a rearranged cellular myc proto-oncogene. Blood. 1986;67:1542-1549

250. Otsuki T, Nakazawa N, Taniwaki M, Yamada O, Sakaguchi H, Wada H, Yawata Y, Ueki A. Establishment of a new human myeloma cell line, KMS-18, having t(4;14)(p16.3;q32.3) derived from a case phenotypically transformed from Ig A-lambda to BJP-lambda, and associated with hyperammonemia. Int J Oncol. 1998;12:545-552

251. Katagiri S, Yonezawa T, Kuyama J, Kanayama Y, Nishida K, Abe T, Tamaki T, Ohnishi M, Tarui S. Two distinct human myeloma cell lines originating from one patient with myeloma. Int J Cancer. 1985;36:241-246

252. Kaufmann H, Ackermann J, Baldia C, Nosslinger T, Wieser R, Seidl S, Sagaster V, Gisslinger H, Jager U, Pfeilstocker M, Zielinski C, Drach J. Both IGH translocations and chromosome 13q deletions are early events in monoclonal gammopathy of undetermined significance and do not evolve during transition to multiple myeloma. Leukemia. 2004;18:1879-1882

253. Tajima E, Uranishi M, Iida S, Komatsu H, Nitta M, Ueda R. Global real-time quantification/reverse transcription-polymerase chain reaction for detecting proto-oncogenes associated with 14q32 chromosomal translocation in multiple myeloma. Haematologica. 2005;90:559-562

254. Chang H, Stewart AK, Qi XY, Li ZH, Yi QL, Trudel S. Immunohistochemistry accurately predicts FGFR3 aberrant expression and t(4;14) in multiple myeloma. Blood. 2005

255. Choi HH, Jong HS, Park JH, Choi S, Lee JW, Kim TY, Otsuki T, Namba M, Bang YJ. A novel ginseng saponin metabolite induces apoptosis and down-regulates fibroblast growth factor receptor 3 in myeloma cells. Int J Oncol. 2003;23:1087-1093

256. Grand EK, Chase AJ, Heath C, Rahemtulla A, Cross NC. Targeting FGFR3 in multiple myeloma: inhibition of t(4;14)-positive cells by SU5402 and PD173074. Leukemia. 2004;18:962-966

257. Paterson JL, Li Z, Wen XY, Masih-Khan E, Chang H, Pollett JB, Trudel S, Stewart AK. Preclinical studies of fibroblast growth factor receptor 3 as a therapeutic target in multiple myeloma. Br J Haematol. 2004;124:595-603

258. Trudel S, Ely S, Farooqi Y, Affer M, Robbiani DF, Chesi M, Bergsagel PL. Inhibition of fibroblast growth factor receptor 3 induces differentiation and apoptosis in t(4;14) myeloma. Blood. 2004

259. Qian S, Somlo G, Zhou B, Zhu L, Mi S, Mo X, Cheung EM, Qiu W, Lin RJ, Rossi J, Holtz M, Chu P, Yen Y. Ribozyme cleavage leads to decreased expression of fibroblast growth factor receptor 3 in human multiple myeloma cells, which is associated with apoptosis and downregulation of vascular endothelial growth factor. Oligonucleotides. 2005;15:1-11

260. Zhu L, Somlo G, Zhou B, Shao J, Bedell V, Slovak ML, Liu X, Luo J, Yen Y. Fibroblast growth factor receptor 3 inhibition by short hairpin RNAs leads to apoptosis in multiple myeloma. Mol Cancer Ther. 2005;4:787-798

261. Lever MA, Th'ng JP, Sun X, Hendzel MJ. Rapid exchange of histone H1.1 on chromatin in living human cells. Nature. 2000;408:873-876

262. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita Ha HA, Kinoshita A, Mizuguchi T, Yoshiura Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N. Haploinsufficiency of NSD1 causes Sotos syndrome. Nat Genet. 2002;30:365-366

263. Smadja NV, Fruchart C, Isnard F, Louvet C, Dutel JL, Cheron N, Grange MJ, Monconduit M, Bastard C. Chromosomal analysis in multiple myeloma: cytogenetic evidence of two different diseases. Leukemia. 1998;12:960-969

264. Smadja NV, Louvet C, Isnard F, Dutel JL, Grange MJ, Varette C, Krulik M. Cytogenetic study in multiple myeloma at diagnosis: comparison of two techniques. Br J Haematol. 1995;90:619-624

265. Barlogie B, Jr., Shaughnessy JD. Early results of total therapy II in multiple myeloma: implications of cytogenetics and FISH. Int J Hematol. 2002;76 Suppl 1:337-339

266. Fassas AB, Spencer T, Sawyer J, Zangari M, Lee CK, Anaissie E, Muwalla F, Morris C, Barlogie B, Tricot G. Both hypodiploidy and deletion of chromosome 13 independently confer poor prognosis in multiple myeloma. Br J Haematol. 2002;118:1041-1047

267. Shaughnessy J, Barlogie B. Chromosome 13 deletion in myeloma. Curr Top Microbiol Immunol. 1999;246:199-203

268. Shaughnessy J, Jr., Tian E, Sawyer J, McCoy J, Tricot G, Jacobson J, Anaissie E, Zangari M, Fassas A, Muwalla F, Morris C, Barlogie B. Prognostic impact of cytogenetic and interphase fluorescence in situ hybridization-defined chromosome 13 deletion in multiple myeloma: early results of total therapy II. Br J Haematol. 2003;120:44-52

269. Zhan F, Barlogie B, Huang Y, Rasmussen E, Sawyer J, Santra M, Walker R, Tricot G, Crowley J, Anaissie E, Shaughnessy J. The Transcriptome of Multiple Myeloma (MM) Defines Disease Subgroups with Distinct Genetic and Clinical Features. Blood. 2004;104

270. Mattioli M, Agnelli L, Fabris S, Baldini L, Morabito F, Bicciato S, Verdelli D, Intini D, Nobili L, Cro L, Pruneri G, Callea V, Stelitano C, Maiolo AT, Lombardi L, Neri A. Gene expression profiling of plasma cell dyscrasias reveals molecular patterns associated with distinct IGH translocations in multiple myeloma. Oncogene. 2005;24:2461-2473

Appendix I

Detailed Screening Data From the t(4;14) Positive Patients

Patient	Diagnosis	Prior Dx	Breakpoint	JH Hybrid	Iu Hybrid	Nested Iu Hybrid	FGFR3 Expression	der(14) Result	Clinical Isotype	Light Chain
434	Smoldering		MB4-3	Detected	Detected		Detected	Positive, IgG	IgG	Lambda
657	MM		MB4-3	Detected	Detected		Detected	Positive, IgA	IgA	Kappa
773	MM		MB4-1	Detected	Detected		Detected	Positive, IgG	IgA	Kappa
830	Smoldering	MGUS	MB4-1	Detected	Detected		Not Detectable	Negative	IgG	Kappa
870	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	
898	MM		MB4-2	Detected	Detected		Not Detectable	Negative	IgG	Lambda
938	MM		MB4-1	Detected	Detected		Detected	Negative	IgG	Kappa
945	MM		MB4-1	Detected	Detected		Detected	Positive, IgG	IgG	Kappa
964	MM		MB4-3	Detected	Detected		Not Detectable	Negative	IgG	Lambda
1022	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Kappa
1039	MM		MB4-1	Detected	Detected	Detected	Detected	Negative	IgG	Lambda
1069	MM		MB4-1	Detected	Detected		Detected	Negative	Kappa	Kappa
1075	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Kappa

Patient	Diagnosis	Prior Dx	Breakpoint	JH Hybrid	Iu Hybrid	Nested Iu Hybrid	FGFR3 Expression	der(14) Result	Clinical Isotype	Light Chain
1076	MM		MB4-1	Detected	Detected		Not Detectable	Negative	IgG	Kappa
1091	MM	MGUS	MB4-1	Detected	Detected		Detected	Positive, IgG	IgG	Kappa
1110	MM		MB4-2	Detected	Detected		Detected	Positive, IgA	IgA	Kappa
1114	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Lambda
1128	MM		MB4-1	Detected	Detected		Detected	Negative	IgG	Kappa
1171	MM		MB4-1	Detected	Detected		Detected	Negative	IgG	Kappa
1174	MM		MB4-2	Detected	Detected		Detected	Positive, IgG	IgG	
1183	MM		MB4-1	Detected	Detected		Detected	Positive, IgG	IgG	Lambda
1193	MM		MB4-1	Detected	Detected		Detected	Negative	IgG	Kappa
1194	MM	Smoldering	MB4-1	Detected	Detected		Not Detectable	Negative	IgG	Lambda
1207	MM		MB4-3	Detected	Not Detected		Not Detectable	Negative	IgG	Kappa
1211	Smoldering		MB4-1	Detected	Detected		Detected	Negative	IgG	Kappa
1223	MM		MB4-2	Detected	Detected		Detected	Negative	IgG	Kappa

Patient	Diagnosis	Prior Dx	Breakpoint	JH Hybrid	Iu Hybrid	Nested Iu Hybrid	FGFR3 Expression	der(14) Result	Clinical Isotype	Light Chain
1237	Smoldering		MB4-1	Detected	Detected	Detected	Detected	Positive, IgG	IgG	Kappa
1244	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Kappa
1264	MM		MB4-2	Detected	Detected		Not Detectable	Negative	Kappa	Kappa
1308	MM		MB4-3	Detected	Not Detected	Not Detected	Not Detectable	Negative	IgA	Lambda
1309	MM		MB4-1	Detected	Detected		Detected	Negative	IgG	Kappa
1332	MM		MB4-1	Detected	Detected		Not Detectable	Negative	IgG	Kappa
1394	MM		MB4-3	Detected	Detected		Detected	Positive, IgA	IgG	Kappa
1461	MM		MB4-1	Detected	Detected		Detected	Negative	IgG	Lambda
1475	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Lambda
1489	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Lambda
1504	MM		MB4-1	Detected	Detected		Not Detectable	Negative	Kappa	Kappa
1510	Smoldering	Smoldering	MB4-1	Detected	Detected	Detected	Detected	Negative	IgG	Kappa
1542	MM		MB4-2	Detected	Detected	Detected	Detected	Negative	IgG	Lambda

Patient	Diagnosis	Prior Dx	Breakpoint	JH Hybrid	Iu Hybrid	Nested Iu Hybrid	FGFR3 Expression	der(14) Result	Clinical Isotype	Light Chain
1560	MM		MB4-1	Detected	Detected		Not Detectable	Negative	Kappa	Kappa
1607	MM		MB4-1	Detected	Not Detected		Not Detectable	Negative	Lambda	Lambda
1631	MM		MB4-1	Detected	Detected	Detected	Not Detectable	Negative	IgG	Lambda
1649	MM		MB4-1	Detected	Detected		Detected	Negative	IgA	Lambda
1661	MM		MB4-3	Detected	Detected	Detected	Detected	Positive, IgG	IgG	Kappa
1307	MGUS		MB4-1	Not Detected	Not Detected	Detected	Detected	Positive, IgG		
1653	MGUS		MB4-1	Detected	Detected	Detected	Not Detectable	Negative	IgG	Lambda

Recently Identified t(4;14) Positive Patients – Not Included in Chapters 1 or 2

Patient	Diagnosis	Prior Dx	Breakpoint	JH Hybrid	Iu Hybrid	Nested Iu Hybrid	FGFR3 Expression	der(14) Result	Clinical Isotype	Light Chain
1704	???		MB4-3	Detected	Detected	Detected	Detected	Negative	IgG	Kappa
1708	MGUS		MB4-3	Detected	Detected	Detected	Not Detectable		IgM +IgG	Lambda
1759	MM		MB4-1	Detected	Detected	Detected	Detected		IgG	
1850	MM		MB4-2	Detected	Detected	Detected	Detected	Positive, IgA	Lambda	Lambda
1856	MGUS		MB4-1	Detected	Detected	Detected	Detected	Negative	IgA	Lambda